

THE F1 CHAPERONE:USHER TRANSLOCON
OF *Yersinia pestis* AND POTENTIAL
APPLICATIONS

by

Noratikah Othman

Thesis submitted in part fulfilment of the requirement for degree of
DOCTOR OF PHILOSOPHY

School of Biological Sciences
The University of Reading
January 2017

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Noratikah Othman

Abstract

Plague is a notifiable disease caused by *Yersinia pestis* for which there is still no reliable vaccine available. The fraction 1 (F1) antigen forms an immunogenic capsule around the cell surface that decreases efficiency of phagocytosis. Many *E. coli* strains and other Gram-negative bacteria coat themselves in similar, remarkably stable protein fibres assembled via specific, dedicated chaperone-ushe (CU) pathways. Fibres are assembled through a pore within each usher protein, via a cycling of specific chaperone:subunit:usher interactions leading to subunit polymerisation and translocation to the cell surface. This project has used bioinformatics analysis and modelling to improve understanding of details of the *Y. pestis* F1 chaperone:usher translocon and has applied this knowledge to enhance export of heterologous epitopes. The number of members of the $\gamma 3$ family of ushers, which assemble simple fibres of one or two subunits, has been expanded and includes for the first time a CU system more closely related to that of the *Yersinia* Caf system. This plasmid encoded usher from the commensal *E. coli* SE11 shares 70% identity with Caf1A. Caf1A usher was modelled using Intfold and ITASSER, based on the crystal structures of *E. coli* FimD and PapC usher in the closed state (PDB: 2VQI, 3.2Å). The complete Caf translocon was modelled based on the open FimD structure (PDB: 3RFZ, 2.8Å) with the Caf1M chaperone:Caf1:Caf1 subunit structure (PDB:1Z9S, 2.2Å) docked in the translocon. Based on the modelled Caf translocon and multiple sequence alignments of related ushers, conserved residues within the β -barrel facing the pore cavity or subunit; residues interacting between the plug and β -barrel in the closed pore and residues interacting between β -barrel and subunit in the 'open' translocating pores were identified. These were then tested by mutagenesis for impact on F1 assembly. No single point mutation within the usher abolished F1 assembly. The most profound effect was following mutation of Ser289 to Ala, the level of surface F1 decreased to $70.81\% \pm 3.0$ of levels with wildtype Caf1A. The potential of F1 fibre to act as a carrier of epitopes was tested by replacing 4 loops within Caf1 with either Gly residues or charged residues. The model was used to understand and optimise limitations in export of modified fibres. One permissive site (loop 5) was identified as the optimum site for loop replacement and surface assembly of modified F1 was enhanced by mutation of a clashing Gln in the β -barrel wall. In addition, the study identified a critical role for a conserved Asn, Asn80, within loop 2 of Caf1 subunit that was modelled interacting with Tyr266 in the plug. The permissive loop 5 was also adapted to export a polyhistidine sequence. While 4 His residues were readily accommodated, F1::1His6 conferred toxicity, although there was still evidence of surface polymer. Thus, this study demonstrated flexibility of the F1 CU pathway for surface display of short peptides and the ability to use the translocon model to identify problem areas and repair export. As F1 is considered to be unique to *Y. pestis*, the presence of a Caf related CU locus in a commensal *E. coli* was of interest to understanding its phylogenetic relationship to F1 and also in diagnostic applications in plague detection. Bioinformatic analysis and recombinant expression of the *E. coli* SE11-P2 CU locus, combined with mass spectrometry analysis, confirmed surface assembly of polymers of both subunits from this locus. A purification strategy for isolation of recombinant SE11-P2 polymers was developed. Purified protein will be used in future studies to raise antibody for further study of this CU system and to test for any cross-reaction with *Y. pestis* F1 antigen, that might potentially interfere with serological tests for *Y. pestis*. F1 antigen remains a component of interest in both anti-plague vaccine design and plague detection. Results from this study have potential to improve reliability of both approaches to plague control.

Acknowledgements

I deeply thank my supervisor *Dr. Sheila MacIntyre* who was very helpful and offered support and guidance during this study. I would also like to thank my co-supervisor *Dr. Liam McGuffin* for his invaluable assistance. To Alma Lopez and Elham Mohebi, thank you for the pACYCF1 construct and loop insertion constructs.

I would like to thank all my *friends*, especially *Norhayati Ngah* who I shared laughter and tears throughout this journey.

A special appreciation to my beloved husband, *Asfar Harris* for his continued and unconditional love and support during the hardest time of my PhD journey.

Finally, to my *mom* and my late *dad*, this is for you both.

Noratikah Othman
University of Reading
Jan 2017

Abbreviations

CU - Chaperone/usher

IAA - iodoacetamide

DSC - Donor strand complementation

DSE - Donor strand exchange

DTT - dithiothreitol

EDTA - ethylenediaminetetraacetic acid

F1 - Fraction 1 antigen

FGL - Long F1G1 loop

FGS - Short F1G1 loop

FTIC - fluorescein isothiocyanate

HMW - high molecular weight

IPTG - isopropyl-D-thiogalactopyranoside

LB - Luria Bertani broth

MS - Mass spectrometry

MSA - Multiple sequence alignment

MWt – molecular weight

OM - Outer membrane

ON - overnight

PAGE - polyacrylamide gel electrophoresis

PBS - phosphate buffered saline

PCR - polymerase chain reaction

PDB - Protein Data Bank

pSE11-2 – Plasmid 2 of *E. coli* SE11

SDS - sodium dodecyl sulphate

Tris - tris(hydroxymethyl) aminoethane

WT – wildtype

PapC_N – PapC N-terminus

FimD_N – FimD N-terminus

Amino acid name	1 letter abbreviation	3 letter abbreviation
Alanine	A	Ala
Arginine	R	Arg
Asparagine	N	Asn
Aspartic acid	D	Asp
Cystiene	C	Cys
Glutamic acid	E	Glu
Glutamine	Q	Gln
Glycine	G	Gly
Histidine	H	His
Isoleucine	I	Iso
Leucine	L	Leu
Lysine	K	Lys
Methionine	M	Met
Phenylalanine	F	Phe
Proline	P	Pro
Serine	S	Ser
Threonine	T	Thr
Tryptophan	W	Trp
Tyrosine	Y	Tyr
Valine	V	Val

Contents

Declaration	2
Acknowledgements.....	4
Abbreviations	5
Contents.....	7
List of Figures.....	14
List of Tables	18
 CHAPTER 1 INTRODUCTION	 20
1.1 <i>Yersinia pestis</i>	21
1.1.1 Overview of <i>Yersinia pestis</i>	21
1.1.2 Plague	22
1.1.3 Virulence Factors of <i>Y. pestis</i>	24
1.2 Significance of F1 antigen in diagnosis and vaccination	27
1.3 F1 capsular antigen.....	31
1.4 Phylogeny of chaperone-usher (CU) families	34
1.5 Assembly of F1 capsular antigen via chaperone-usher pathway	41
1.5.1 Donor strand complementation (DSC) of subunit by chaperone.....	42
1.5.2 Donor strand exchange (DSE) at periplasmic surface of Caf1A usher....	43
1.6 Assembly at the <i>E. coli</i> ushers PapC and FimD	46
1.7 Aims and objectives	52

CHAPTER 2 MATERIAL AND METHODS	54
2.1 Bacteria strain	55
2.2 Media and culture conditions.....	55
2.3 Plasmids and mutants	56
2.4 DNA techniques.....	61
2.4.1 Plasmid isolation.....	61
2.4.2 DNA quantitation.....	62
2.4.3 Agarose gel electrophoresis.....	62
2.4.4 DNA sequencing and analysis	63
2.5 Mutagenesis and Polymerase Chain Reaction (PCR)	63
2.5.1 Oligonucleotides	63
2.5.2 Site-directed mutagenesis.....	71
2.5.3 Deletion mutagenesis	72
2.5.4 InFusion Mutagenesis	73
2.5.5 DpnI digestion.....	77
2.5.6 Purification of PCR products.....	77
2.5.7 DNA extraction from agarose gels.....	78
2.6 Making competent cells.....	78
2.7 Transformation.....	79
2.8 Cell fractionation.....	80
2.8.1 Analysis of F1 in whole <i>E. coli</i> cells.....	80
2.8.2 Isolation of surface F1	80
2.8.3 Isolation of periplasmic F1	81
2.9 Protein techniques	82
2.9.1 Bradford assay	82
2.9.2 Quantitation of Caf1 from Coomassie blue stained SDS-PAGE.....	83

2.9.3	Sodium-dodecyl-sulphate polyacrylamide gel electrophoresis	84
2.9.4	Immunoblotting.....	84
2.9.5	TCA precipitation	85
2.9.6	Protein identification by Mass spectrometry	86
2.9.7	Purification of <i>E. coli</i> SE11 subunits	87
2.10	Bioinformatics analysis.....	89
2.10.1	Basic Local Alignment Similarity Search (BLAST)	89
2.10.2	CLUSTALW 2.0.....	89
2.10.3	PHYLIP 3.695	89
2.10.4	SignalP.....	90
2.10.5	BetaCavityWeb	91
2.10.6	MEGA 5.2	91
2.10.7	I-TASSER.....	91
2.10.8	IntFOLD2	91
2.10.9	TMalign	92
2.10.10	PyMOL version 1.3.....	92
2.10.11	DiscoTope 2.0.....	93

CHAPTER 3 ANALYSIS OF F1 EXPORT IN NEW SYSTEM – pACYCF1..... 94

3.1	Introduction.....	95
3.2	Comparison of assembled F1 and chaperone from pACYCF1 and pFMA1 ...	96
3.3	Optimisation of induction time and Caf1 extraction temperature	98
3.3.1	Thermoinduction of pACYF1.....	98
3.3.2	Optimisation of F1 heat extraction for quantitation.....	100
3.3.3	Comparison of Caf1 from whole cells, surface extracted and TCA precipitation of culture media	101
3.4	Caf1 quantitation following SDS-PAGE.....	102

3.5	Conclusions.....	105
-----	------------------	-----

CHAPTER 4 MODELLING AND PROPERTIES OF THE CAF TRANSLOCON **107**

4.1	Introduction.....	108
4.2	Extension of $\gamma 3$ family of fimbrial usher protein.....	109
4.2.1	Structural proteins loci of the $\gamma 3$ family members.....	114
4.3	Operon analysis and domain analysis of <i>E. coli</i> SE11 caf like genes.....	115
4.4	Modelling of the Caf translocon	118
4.4.1	Modelling of closed and opened Caf1A.....	119
4.4.2	Modelling of <i>E. coli</i> SE11 translocon	125
4.5	Characterisation of Caf barrel.....	126
4.5.1	Residues interacting with conserved residues in the plug domain	126
4.5.2	Localisation of key residues within Caf1A barrel	129
4.6	Characterisation of important residues in Caf1A by mutagenesis	134
4.6.1	Investigating the role of Tyr266 and Glu298	135
4.6.2	Mutation of residues predicted to be involved in pore gating	137
4.6.3	Investigating the role of conserved residues in β -barrel	140
4.7	Conclusions.....	142

CHAPTER 5 CAF1 AS A SCAFFOLD FOR EXPORT OF HETEROLOGOUS EPITOPES..... **145**

5.1	Introduction.....	146
5.2	Prediction of B-cells epitopes in Caf1	147
5.3	Generation and phenotype of constructs with glycine and charge epitope inserts.....	149
5.3.1	Effect of F1 loop mutations on bacterial cell phenotype, initial screen .	154
5.3.2	Impact of loop mutations on surface assembly of Caf1	156

5.3.3	Impact of loop mutations on stability of the F1 polymer	158
5.3.4	Quantitative comparison of mutated F1 polymers on the cell surface ...	160
5.4	In silico analysis of F1 loop insertions	164
5.5	Increasing surface production of F1 with loop insertions by site-directed mutagenesis	171
5.5.1	Attempt to repair clashes from loop 1 in the F1::3 construct	173
5.5.2	Replacement of required Asn residue in loop 4 of pACYCF1D12 and pACYCF1::2 constructs.....	175
5.5.3	Repairing clash and new interaction involving loop 5 of F1D13 and F1::1	178
5.5.4	Repairing loss of interaction in loop 7 of pACYCF1D14 and pACYCF1::4 constructs.....	181
5.6	Histidine insertion in loop 1	182
5.7	Conclusions.....	185
CHAPTER 6 CHARACTERISATION OF E. COLI SE11.....		187
6.1	Introduction.....	188
6.1.1	Comparison of <i>Y. pestis</i> pMT1/pFRA plasmid with commensal <i>E. coli</i> pSE11-2.....	189
6.1.2	Pairwise alignment with Caf1 and multiple sequence alignment of subunits	191
6.1.3	Homology modelling of <i>E. coli</i> SE11 subunits and identification of conserved residues compared to Caf1	194
6.1.4	Prediction of B cell epitopes in <i>E. coli</i> SE11 fimbrial subunits	195
6.2	Construction and characterisation of recombinant <i>E. coli</i> SE11 CU operon	198
6.2.1	Cloning <i>E. coli</i> SE11 caf-like operon for recombinant fibre assembly and tagged chaperone-subunit complex	198
6.2.2	Lack of toxicity of cloned CU operon and identification of expressed protein	201
6.2.3	Comparison of recombinant production of periplasmic chaperone-subunit complexes and surface assembled subunit from the <i>E. coli</i> SE11 fimbriae operon... ..	203

6.2.4	Heat stability of recombinant surface assembled polymer	206
6.3	Mass spectrometry of <i>E. coli</i> SE11 subunit.....	207
6.4	Purification of recombinant subunits from <i>E. coli</i> SE11.....	209
6.4.1	Comparison of His tagged periplasmic chaperone complexes	209
6.4.2	Purification of <i>E. coli</i> SE11 chaperone:subunit/s complexes by Ni ²⁺ NTA.	210
6.4.3	Purification of recombinant surface assembled <i>E. coli</i> SE11 subunits by size exclusion chromatography	211
6.5	Conclusions.....	214
Chapter 7 GENERAL DISCUSSION AND FUTURE DIRECTIONS		216
7.1	Properties of the Caf1A usher.....	217
7.2	Caf1 as scaffold of short peptides	218
7.3	<i>E. coli</i> SE11 fimbriae operon similar to caf operon.....	218
7.4	Future directions	219
Appendix 1: Usher proteins from similarity search results, using BLASTP, used to construct phylogenetic tree.		221
Appendix 2: Multiple sequence alignment of all γ 3 fimbrial.		225
Appendix 3: Impact on subunit polymerization using plasmid pFM1 and surface polymer using plasmid pFMA1.		227
Appendix 4: Caf1 sequence with G or ELDKWA insertion in FASTA format. Coloured in red and italic indicates replacement		228
Appendix 5: Alignment of pACYC-F1D13 sequencing results.		229
Appendix 6: Alignment of pACYC-F1D11 sequencing results		230
Appendix 7: Alignment of pACYC-F1D12 sequencing results.		231
Appendix 8: Alignment of pACYC-F1D14 sequencing results.		232
Appendix 9: Alignment of pACYC-F1::1 sequencing results.....		233
Appendix 10: Alignment of pACYC-F1::2 sequencing results.		234

Appendix 11: Alignment of pACYC-F1::3 sequencing results.	235
Appendix 12: Alignment of pACYC-F1::4 sequencing results.	236
Appendix 13: Whole operon sequencing of pACYC-F1D12, F1::2, F1::3 and F1::4 constructs	237
Appendix 14: Model of protein in this study attached in a CD	239
Appendix 15: SDS-PAGE of F1 from cells carrying polyhistidine insertion at loop 5	240
References.....	241

List of Figures

Figure 1.1: Architecture of F1 fibre..	33
Figure 1.2: Assembly of CU pili and curli is driven by simple systems, involving specific periplasmic chaperones and outer membrane ushers.	38
Figure 1.3: Phylogenetic tree of the fimbrial usher family.	40
Figure 1.4: Summary of Caf CU pathway.	41
Figure 1.5: Topology Diagram and sequence of Caf1A.	42
Figure 1.6: Caf1M:Caf1 complex allows proline lock (Pro104m) to open.	44
Figure 1.7: Topology diagram of Caf1M:Caf1:Caf1 interactions.	45
Figure 1.8: Crystal structure (3.2Å resolution) of PapC, PDB: 2VQI.	48
Figure 1.9: Closed and open structures of FimH.	50
Figure 2.1: In-Fusion cloning procedure using In-Fusion HD Cloning kit (Image taken from Clonotech In-Fusion user manual)	76
Figure 3.1: Comparison of pFMA1 and pACYCF1 plasmid.	96
Figure 3.2: SDS-PAGE analysis to compare level of surface F1 in <i>E. coli</i> DH5α carrying either pACYCF1 or pFMA1.	97
Figure 3.3: Comparison of level of Caf1M chaperone in induced cultures of <i>E. coli</i> DH5α carrying pACYCF1 or pFMA1.	98
Figure 3.4: Analysis of F1 production in six transformants of <i>E. coli</i> DH5α/pACYCF1 expression from whole cells.	99
Figure 3.5: SDS-PAGE analysis of thermoregulation of F1 at 26°C, 28°C and 37°C.	100

Figure 3.6: Impact of extraction temperature on recovery of surface F1. F1 polymer was extracted from cells of <i>E. coli</i> DH5 α /pACYCF1, thermoinduced for 4 hour, by incubation at 57°C or 65°C for 1 hour	101
Figure 3.7: SDS-PAGE analysis of F1 in whole cells, surface extraction and TCA precipitate from culture media.	102
Figure 3.8: Standard gel for F1 quantitation.	104
Figure 3.9: F1 standard concentration graph for determination of F1 standard reliability.	105
Figure 4.1: Phylogenetic tree of Gamma fimbriae group of usher proteins.	112
Figure 4.2: Phylogenetic tree of fimbrial usher protein family.	113
Figure 4.3: <i>E. coli</i> SE11 plasmid encoded chaperone-usher cluster (pSE11-2 (NC_011413.1)).	116
Figure 4.4: Quality and reliability assessment of Caf1A models based on ModFOLD4.	120
Figure 4.5: Measurement of Caf1A pore modelled on FimD.	123
Figure 4.6: Model of assembled Caf translocon.	125
Figure 4.7: Superposition of modelled <i>E. coli</i> SE11 usher (pink) and Caf1A (green) with structure of FimD (blue).	126
Figure 4.8: Multiple sequence alignment and interaction of middle domain (plug) with β -barrel and subunit.	128
Figure 4.9: Usher cavity predicted by BetaCavityWeb.	131
Figure 4.10: Analysis of important residues in Caf1A β -barrel.	133
Figure 4.11: SDS-PAGE of surface F1 extracted from <i>E. coli</i> BW25113 from pACYCF1_Y266A and pACYCF1_Y266S constructs.	137
Figure 4.12: Superpose of closed PapC structure (3FIP, white) and Caf1A closed model (yellow).	137
Figure 4.13: Superpose of closed PapC structure (3FIP, white) and Caf1A closed model (yellow).	140
Figure 4.14: SDS-PAGE of extracted surface F1 from <i>E. coli</i> /pACYCF1 with point mutations in Caf1A.	141
Figure 5.1: Location of Gly and ELDKWA replacements within Caf1 sequence.	148
Figure 5.2: Epitope prediction in F1.	148

Figure 5.3: Caf1 ₂ :Caf1M complex (PDB: 1Z9S, 2.2 Å) showing surface loops tested for epitope replacement.....	149
Figure 5.4: Plasmid source of mutated loop sequences.....	150
Figure 5.5: Workflow of In-fusion HD protocol.....	152
Figure 5.6: Electrophoresis analysis of DpnI digested and purified linearised vector, and purified inserts.....	153
Figure 5.7: Comparison of F1 production from <i>E. coli</i> Top10 carrying pACYCF1 encoding different loop mutations.....	155
Figure 5.8: SDS-PAGE of surface F1 with loop replacements.	157
Figure 5.9: Temperature stability	159
Figure 5.10: Western blot of temperature stability of extracted F1 from loop mutation constructs.	159
Figure 5.11: Quantitation of surface F1 extracted from <i>E. coli</i> BW25113 expressing F1 with loop mutations (pACYC-F1D11, -F1D12, -F1D13, -F1D14 and -F1::1).	162
Figure 5.12: Quantitation of surface F1 extracted from <i>E. coli</i> BW25113 expressing F1 with loop mutations (pACYC-F1::2, -F1::3 and -F1::4).	163
Figure 5.13: Superimposed Caf1 mutated model with Caf1 wild type structure to compare the size of insertion in each loops.	165
Figure 5.14: Translocon of loop mutation.	170
Figure 5.15: SDS-PAGE of extracted surface F1 from <i>E. coli</i> /pACYCF1 with point mutations in Caf1A.	174
Figure 5.16: SDS-PAGE of surface F1 extracted from <i>E. coli</i> BW25113 from construct pACYCF1::2_Y266A and pACYCF1::2_Y266S.	175
Figure 5.17: SDS-PAGE of surface F1 extracted from <i>E. coli</i> BW25113 to assess and improve export of loop 4 mutation.....	177
Figure 5.18: SDS-PAGE of extracted surface F1 from <i>E. coli</i> /pACYCF1 with point mutations in Caf1A.	177
Figure 5.19: SDS-PAGE of extracted surface F1 from <i>E. coli</i> /pACYCF1 with point mutations in Caf1A.	180
Figure 5.20: SDS-PAGE of surface F1 containing F1::4_N81G.....	181
Figure 5.21: SDS-PAGE of surface F1::1DS6His.	183
Figure 5.22: Growth curve of pACYCF1_DS6his_Q169A and pACYCF1_DS4his. .	184

Figure 6.1: Comparison of plasmids.	190
Figure 6.2: Multiple sequence alignment (ClustalW) and pairwise sequence alignment (EMBOSS) of <i>E. coli</i> SE11 subunits with Caf1.	192
Figure 6.3: Prediction of signal peptide cleavage using SignalP.	193
Figure 6.4: Modelled <i>E.coli</i> SE11 fibre subunits.	194
Figure 6.5: B cell epitope prediction using DiscoTope.	197
Figure 6.6: pACYCD1-MA-S123 and pTRC99-MHis-S123 constructs.	200
Figure 6.7: Growth curve of recombinant <i>E. coli</i> BL21 (DE3)/pACYCD1-MA-S123	202
Figure 6.8: SDS-PAGE of surface and periplasmic subunits/polymer of <i>E. coli</i> SE11 fimbriae operon.	205
Figure 6.9: Temperature stability of surface extracted cells carrying pACYCD1-MA-S123.	206
Figure 6.10: SDS-PAGE of surface extracted <i>E. coli</i> SE11 fimbriae.	208
Figure 6.11: SDS-PAGE of periplasmic subunits/polymer from pTRC99-Mhis-S1 and pTRC99-M8His-S123.	210
Figure 6.12: NTA purification of recombinant <i>E. coli</i> SE11 chaperone:subunits complexes.	212
Figure 6.13: Sephacryl S300 purification of <i>E. coli</i> SE11 subunits/polymer. Chromatogram and SDS-PAGE of pACYCD1-MA-S123.	213

List of Tables

Table 1.1: Virulence factors of <i>Yersinia pestis</i> (adapted from (Butler, 2009))	26
Table 1.2: Function-structure classification of chaperone-usher fimbrial adhesins.	36
Table 2.1 List of competent cells used in this study	55
Table 2.2: Plasmids used in this study	57
Table 2.3: Primers used in the construction of mutants.	64
Table 2.4: Sequencing primers	70
Table 2.5: PCR reaction mixture of inverse PCR	72
Table 2.6: PCR reaction condition for inverse PCR.	73
Table 2.7: Preparation of RF1 and RF2 buffer.	79
Table 2.8: Preparation of microplate gamma-globulin standard dilutions, sufficient for duplicates..	83
Table 2.9: Preparation of microplate samples of F1.	83
Table 3.1: Estimated concentration of total protein in F1 standard.	103
Table 3.2: The raw data of F1 standard from each lane of SDS PAGE.	104
Table 4.1: Locus structure of updated $\gamma 3$ fimbriae group CU systems.	115
Table 4.2: In silico analysis of <i>E. coli</i> SE11 pSE11-2 chaperone-usher cluster.	117
Table 4.3: Summary of Caf1A residues with potential importance on Caf translocon function.	134
Table 5.1: Caf1 loop mutation in pACYCF1.	150

Table 5.2: Quality and confidence level of each mutated Caf1 model.	166
Table 5.3: List of point mutation and insertion to improve F1 export in loop mutation constructs.	171
Table 6.1: Summary of conserved residues in Caf1 based on MSA of Caf1 and <i>E. coli</i> SE11 subunit 1 and 2.....	195
Table 6.2: Peptides identified by mass spectrometry	208

CHAPTER 1 INTRODUCTION

1.1 *Yersinia pestis*

1.1.1 Overview of *Yersinia pestis*

Yersinia pestis is a gram-negative coccobacillus bacteria. It is non-motile and facultatively anaerobic. It exists as a single organism or in short chains that can infect humans and animals; it causes the plague. Alexandre Yersin and Shibasaburo Kitasato first described *Y. pestis* as the causative agent of plague during the Hong Kong epidemic in 1894. Alexandre Yersin extracted the fluid of enlarged lymph nodes from people who had died of the plague and injected into guinea pigs, which subsequently died. Organs of the guinea pigs were also infected with the same bacteria. He then discovered that the same microorganism infects both rats and humans causing the plague. The bacterium was subsequently named *Yersinia pestis* in honour of Yersin's initial accurate observations and detailed study (Bendiner, 1989; Rollins et al., 2003).

Y. pseudotuberculosis is the most closely related bacterium to *Y. pestis* where they share more than 97% nucleotide identity in approximately 75% of their genes (Achtman et al., 1999; Skurnik et al., 2000). It was estimated that *Y. pestis* emerged from *Y. pseudotuberculosis* as recently as the 14th century (Bos et al., 2011; Parkhill et al., 2001). However, a paper published in 2015 reported that *Y. pestis* causes plague emerged from less pathogenic *Y. pestis* that was discovered from the teeth of Bronze Age humans originated around 5000 years ago. This less pathogenic *Y. pestis* does not cause bubonic plague because the lack of the murine toxin required for growth of this bacteria in flea gut, however, it is able to cause septicemic and pneumonic plague (Rasmussen et al., 2015).

1.1.2 Plague

There have been 3 major pandemics of plague; The Justinian plague, Black Death and the most recent pandemic called Modern plague. The Justinian plague began in the 541 AD. It was spread across trade routes around the Mediterranean Sea and causes death between 50% and 60% of population in North Africa, Europe and central and South Asia (Perry and Fetherston, 1997; Inglesby et al., 2000; Ligon, 2006). The Justinian plague took place in early 1330s, originated from China and spreading to Europe through trade routes. This pandemic killed around 60% of the European population (Inglesby et al., 2000; Slack, 1989). In 1855, the modern plague arose from China and spread to Asia, Africa and America, carried by rats that travelled on steamships, and killed over 13 million people (Darling et al., 2002; Stenseth et al., 2008; Keim and Wagner, 2009). In total, it is estimated that nearly 200 million people were killed in these three great pandemics (Perry and Fetherston, 1997). In the 21st century approximately 4000 plague cases are reported annually worldwide (Stenseth et al., 2008).

People usually contract plague from being bitten by a rodent flea that is carrying *Y. pestis* or by handling the dead rodent. There are many species of fleas that can be the vectors of plague, but the most effective vector is the oriental rat flea, the *Xenopsylla cheopis* (Titball and Williamson, 2001). The initial step in infection is that the flea ingests *Y. pestis* from the rodent host. Then, *Y. pestis* multiplies in the flea gut. The ingested blood clots lay in the proventriculus. The haemin storage locus (*hms*) helps *Y. pestis* to colonise the proventriculus and the flea is unable to move food from its esophagus to the mid gut (Darby et al., 2002). Consequently, the flea eats more aggressively and regurgitates the infected blood into the dermis of the mammal on which it is feeding, in this case human. During this process,

approximately 25,000 to 100,000 *Y. pestis* are inoculated into the skin of the mammal host (Raoult et al., 2013; Rollins et al., 2003).

When a human is bitten by an infected flea or gets infected by handling animal tissues, this causes bubonic plague. This is the most common form of plague. *Y. pestis* then multiplies in the lymphatic system and causes the lymph nodes to swell, known as bubo. The incubation period of bubonic plague is between 2-6 days and if left untreated, it can have a 40% to 70% mortality rate and mortality reduces to 10% with antibiotic treatment (Prentice and Rahalison, 2007; Rollins et al., 2003; Rosenzweig et al., 2011; Stenseth et al., 2008).

Septicemic plague may develop from bubonic plague where the bacteria access the liver and spleen through the blood stream. The bacteria then replicates and spreads to other organs causing septicaemia and death by multiple organ failure (Sun et al., 2011). It also can occur without a bubo where the bacteria replicate in other tissues and enter the blood stream. The symptoms of septicemic plague begin approximately between 1-4 days after infection by *Y. pestis*. The mortality rate can be reduced with immediate antibiotic treatment, but humans infected by this type of plague generally die if left untreated (Stenseth et al., 2008; Rollins et al., 2003; Prentice and Rahalison, 2007).

The third form of plague is pneumonic plague which can be spread through inhalation. *Y. pestis* invades the lungs and results in person to person transmission via air droplets. Pneumonic plague can also develop following initial bubonic or septicemic plague and then spread to the lungs. This form of plague is the most dangerous with nearly 100% mortality and death in 18 to 24 hours if left untreated (Prentice and

Rahalison, 2007; Hatkoff et al., 2012). However, the mortality can be reduced to 15% by streptomycin or gentamycin within 24 to 36 hours of infection (Rosenzweig et al., 2011).

Despite its ability to cause natural infection, *Y. pestis* is also classified as category A biological weapon. In the 14th century, Tatar Forces used plague victims and catapulted them to Kaffar (now known as Feodosia, Ukraine). Then, Japanese forces used plague-infected fleas and released it from aircraft over Chinese cities during World War II and the most recent case of use in bioterrorism was a man from Ohio who was connected to an extremist group and who tried to obtain *Y. pestis* from the American Type Culture collection (Mierzwa, 1975; Atlas, 1998; Sun et al., 2011). *Y. pestis* can be easily obtained from animal reservoirs making it an important bioterrorism threat (Perry and Fetherston, 1997).

1.1.3 Virulence Factors of *Y. pestis*

The first complete *Y. pestis* genome was sequenced in 2001 from *Y. pestis* strain CO92 (Parkhill et al., 2001). This bacterium possesses a 4.65 Mb chromosome and three plasmids of 100 kb, 70.3 kb and 9.6 kb. The 70.3 kb plasmid known as pYV or pCD1 is a common virulence plasmid throughout pathogenic *Yersinia*, therefore it is not species specific. On the other hand, *Y. pestis* has acquired two plasmids (pPCP1 and pMT1) that encode a range of virulence determinants making it unique (Parkhill et al., 2001). The common low calcium response pYV virulence plasmid encodes a type 3 secretion system (T3SS), that forms a needle like injectisome. This structure mediates transport of *Yersinia* outer proteins (Yops) directly into host cells where they modulate the host immune response, blocking phagocytosis, and triggering host cell death (Hatkoff et al., 2012). It also encodes V antigen (LcrV), a major protective immunogen that forms the tip of the needle (Brubaker, 2003). V antigen also upregulates interleukin-10 (IL-10) and at the same time downregulates

tumor necrosis factor-alpha (TNF α) causing anti-inflammatory mechanisms and downregulates interferon-gamma (IFN γ) causing immunosuppression (Brubaker, 2003). Expression of these proteins is temperature regulated, with expression at 37°C with a low calcium concentration. In a study using a rat model, infection with pYV negative strains of *Y. pestis* led to more polymorphonuclear leucocytes (PMNs) with fewer bacteria in bubos compared to wildtype pYV. This supports the earlier evidence that pYV suppresses the PMN response (Chanteau et al., 2003).

The smaller unique plasmid (9.6 kb) is the pPst plasmid, also known as pPCP1. This plasmid encodes the plasminogen activator, Pla which plays a role in dissemination of *Y. pestis* by lysing fibrin clots and releases the bacteria to peripheral organs. This process takes place at 37°C. Pla is also known as a key virulence factor for bubonic and pneumonic plague (Parkhill et al., 2001; Rosenzweig et al., 2011). It also reduces attraction of leukocytes to infected site by cleaving central complement component system (C3) (Lathem et al., 2005). The pH6 antigen (PsaA) expresses from chromosomal locus under regulation of RovA which plays a role as antiphagocytic. It expresses at 37°C in acidic medium, has function similar to F1 except it acts independently of F1 to resist phagocytosis by mouse and human macrophage (Huang and Lindler, 2004). Other virulence factors that are important are the non-fimbrial adhesins from *yadBC* operon that allow bacteria to adhere to epithelial cells and macrophages (Forman et al., 2008), lipopolysaccharide and lipoprotein that acts together to cause septic shock (Sha et al., 2008; Knirel et al., 2006).

The second unique plasmid is known as pFra or pMT1 (Cherepanov et al., 1991; Protsenko et al., 1983; Parkhill et al., 2001). The size of the plasmid is variable between strains. In *Y. pestis* C092 (Parkhill et al., 2001), the size is 96.2 kb and 100.9 kb in *Y. pestis* KIM (Deng et al., 2002). This plasmid encodes two renowned virulence factors which are murine toxin that has phospholipase D activity and fraction 1 (F1) capsular antigen. *Y. pestis*

needs phospholipase D as it helps the bacteria to multiply and survive in flea midgut (Hinnebusch et al., 2002). The F1 capsule has been shown to prevent adhesion-receptor interactions thus reduces the ability of monocytes, macrophages and dendritic cells to take up bacteria cells. The F1 antigen is expressed at 37°C that allows the bacteria to multiply while resisting phagocytosis (Du et al., 2002; Hatkoff et al., 2012).

Table 1.1: Virulence factors of *Yersinia pestis* (adapted from (Butler, 2009))

Product	Role
Type III secretion system (T3SS) (pYV or pCD1)	Insertion of Yops through a needlelike injectisome into neutrophils, macrophages and dendritic cells (Marketon et al., 2005).
V antigen (LcrV) (pYV or pCD1)	Anti-inflammatory activity through interleukin 10 secretion and form tip of type III secretion system (Brubaker, 2003; Mueller et al., 2005).
<i>Yersinia</i> outer proteins (Yops) (pYV or pCD1)	Inhibit phagocytosis, platelet aggregation and cytokine production (Pouliot et al., 2007; Zhou et al., 2005).
Fraction 1 capsular antigen (F1) (pFra or pMT1)	Inhibition of phagocytosis (Du et al., 2002).
Murine toxin (pFra or pMT1)	Survival of bacteria in flea gut (Hinnebusch et al., 2002).
Plasminogen activator (Pla) (pPst or pPCP1)	Activation of host plasminogen to plasmin for lysis of clots; cleavage of extracellular matrix to promote bacterial spread and bacterial multiplication in lung (Lathem et al., 2005).
Hemin storage (Chromosomal pigmentation locus (pgm))	Formation of obstructive biofilm in flea gut (Carniel, 1999).
<i>Yersinia</i> bactin (Chromosomal pigmentation locus (pgm))	Siderophore for iron transport (Huang and Lindler, 2004).
pH6 antigen (Chromosomal locus under regulation of RovA)	Inhibition of phagocytosis (Cathelyn et al., 2006; Huang and Lindler, 2004).
Outer membrane (Chromosomal locus for Ail proteins)	Resistance to complement-mediated killing (Bartra et al., 2008).
Lipopolysaccharide (Chromosomal locus for cell wall synthesis)	Initiation of inflammatory responses leading to septic shock, also for anti-inflammatory no stimulation of Toll-like receptor 4 (Li and Yang, 2008).

1.2 Significance of F1 antigen in diagnosis and vaccination

According to the World Health Organisation (WHO), recent plague outbreak reported in Madagascar in September 2014 resulted in 27% fatality rate (71 deaths) out of 263 cases. In July 2014, a plague case was reported in China where a man died of bubonic plague and 30,000 residents of Yumen city were quarantined from the outside world. In 2013, a plague outbreak was reported in Madagascar with 84 plague cases and 32 deaths. In 2012, there were 256 cases of plague and 60 deaths reported in Madagascar. In 2010, it was confirmed that 17 people were infected with plague in Peru. Out of 17 cases, 4 were pneumonic, 12 were septicemic and one bubonic (World Health Organization, 2013). The emerging of plague cases every year, the potential for rapid spread of the disease through air, and classification of plague as a category A biological weapon means that it remains a big problem and that it is very important to prevent and to treat the disease.

A few diagnostic methods are used to detect *Y. pestis* in clinical specimens. Since the discovery of the bacteria in Hong Kong, the microorganism was isolated and cultured from bubo aspirates, blood or sputum specimen. This method is sensitive to detect the bacteria however it takes 2-3 days for identification. The other method is to gram stain or Wayson stain blood from the bubo aspirate. This diagnostic is rapid but with moderate sensitivity (Butler, 2009). For rapid diagnostic, PCR (Rahalison et al., 2000) and ELISA (Chanteau et al., 2000) tend to be used for detection of *Y. pestis* from bubo aspirates with sensitivity of moderate and high respectively. Real-time PCR is also used for rapid and specific test for plague diagnosis. In one study, bacteraemia in blood culture counts from patient with bubonic plague ranged from 10^4 - 10^7 cfu/mL (Perry and Fetherston, 1997) and real-time PCR with a probe for F1 antigen can detect 10^2 - 10^4 cfu/mL in less than 5

hours, which is very sensitive (Loiez et al., 2003). Probe for detection of 16S RNA can give the results in less than 3 hours (Tomaso et al., 2003). A rapid diagnostic test (RDT) dipstick was developed for use 'at the bedside'. It was trialled in Madagascar and shown to detect F1 antigen at a concentration as low as 0.5 ng/mL (Chanteau et al., 2003). Efficacy of this assay is highlighted by the fact that the concentration of F1 antigen in patients who have died of plague is generally > 1 µg/mL (Chanteau et al., 1998). The RDT dipstick is very sensitive on bubo aspirates and other specimens compared to culture and ELISA assay.

Antibiotic therapy has been used since 1948 where streptomycin was used to save lives in the Vietnam war (1960-1975). However, use of streptomycin was discontinued in most countries in the world, except Madagascar due to nephrotoxicity and ototoxicity (Butler, 2009). Madagascar produced its own streptomycin to treat bubonic plague and streptomycin was used together with trimethoprim-sulfamethoxazole. This combination saved the lives of two boys in 1995 where isolates from the first boy showed that the bacteria were resistance to streptomycin, chloramphenicol, tetracycline, sulphonamide, ampicillin, kanamycin, spectinomycin but sensitive to trimethoprim. The second case was a 14 year old boy with a strain resistance to streptomycin. Gentamycin has been used successfully in treating plague in the United States (Boulanger et al., 2004) and Tanzania (Mwenge et al., 2006). Tetracycline and doxycycline are also effective treatments for plague. They can be taken orally and also used to prevent pneumonic plague (Butler, 2009). In all cases, patients must be treated immediately with antibiotic to prevent fatality.

Vaccines have been used in the past; in the military and for laboratory workers but available vaccines have limitations. Two forms of vaccine have been used, the live attenuated vaccine and the killed whole cell vaccine. In 1946, a killed whole cell vaccine

(USP) was developed for use in humans and F1 may be the primary protective component (Infectious Diseases Society of America. et al.). Vaccines were prepared by heating or treatment with formaldehyde. However, 10% of those immunised with killed whole cell vaccines suffered side effects such as malaise, headache, elevated temperature and lymphadenopathy (Titball and Williamson, 2001). USP was used to immunise US soldiers that served in Vietnam war in 1960s-1970s and was effective to prevent bubonic plague. However, incidences of pneumonic plague were reported at that time (Anderson et al., 1996; Meyer, 1970). There is also evidence from studies with mice and non-human primates that immunisation of killed whole cells protects against subcutaneous challenge but not against inhalation challenge with *Y. pestis* F1⁺ strain (Meyer, 1970; Titball and Williamson, 2004; Meyer et al., 1974a; Cohen and Stockard, 1967). The same challenges were tested on mice with an F1⁻ strain and showed no protection at all (Anderson et al., 1996). This highlights that F1 is the protective antigen, but would suggest that it is the only major protective antigen in the killed whole cells vaccine. Moreover, repeat immunisations of USP were required for protection. For these reasons, since 1999, this vaccine has not been used in the United States (Jones et al., 2003).

A live attenuated vaccine, EV76, which has been used since 1908, is derived from a pigmentation negative derivative (Pgm) of a fully virulent strain of *Y. pestis* (Zhou and Yang, 2009; Eisen et al., 2006). A study from Lindler in 1998 showed that immunisation of mice with the EV76 vaccine protects mice against subcutaneous and inhalation challenges with *Y. pestis*. Despite the protection given by this vaccine, safety concerns of the use of this live vaccine have been raised because it is not avirulent. It caused fatality in approximately 1% of mice that were immunised with this strain (Lindler et al., 1998; Russell et al., 1995), exhibited virulence in non-human primates (Meyer et al., 1974b)

and was reactogenic in humans (Feodorova and Corbel, 2009). Williamson and co-workers concluded that killed whole cells vaccine (USP) protect solely against bubonic plague due to lack of the T3SS V antigen in the formulation, while the live attenuated EV76 vaccine protects against both bubonic and pneumonic plague as both F1 and V antigen are present in the formulation (Williamson et al., 1995).

Because of the poor protection from whole cell vaccines and safety concerns with the live attenuated EV76 vaccine, there has been more recent focus on subunit and recombinant vaccines. The two prime candidates for plague subunit vaccines have been the F1 and V antigens as discussed above. The F1 protein inhibits phagocytosis and produces a strong humoral response to F1, thus it remains a major component of a new generation of subunit vaccines. Purified preparations of both of these antigens have been tested on non-human primates (*Cynomolgus* macaque) and demonstrated to give protection against aerosol challenge when used individually or together (Cornelius et al., 2008; Fellows et al., 2010; Williamson et al., 2011). The combined F1 and V antigen gave stronger protection than either individual antigen and the killed and attenuated whole cell vaccines (Titball and Williamson, 2001; Williamson et al., 1995). An alternate F1-V vaccine created by genetic fusion of *cafI* and *lcrV* was also tested on macaques and baboons (Stacy et al., 2008) and was shown to give protection against bubonic plague. However, recombinant F1-V vaccine did not protect African green vervets despite its strong antibody response (Mizel et al., 2009). While the F1 sequence is identical in all strains, limited polymorphism of *lcrV* in *Y. pestis* has been reported. This together with concern over F1⁻ strains has continued a search for additional protective antigens (Feodorova and Corbel, 2009; Feodorova and Motin, 2012).

There are many other new approaches to developing vaccines, particularly using heterologous bacteria, viral and plant delivery systems (Feodorova and Motin, 2012). The best characterised bacteria that has been used as a live carrier platform are strains of *Salmonella*. Attenuated *Salmonella enteritica* serovars Minnesota, *Typhimurium*, and *Typhi* have all been used to express F1, LcrV and a combination of both antigens (Feodorova and Motin, 2012; Andrews et al., 1996; Motin et al., 1994). This system also has been tested to express other surface antigens such as HmuR and PsaA (pH6 antigen) (Feodorova and Motin, 2012). *Salmonella* like *Yersinia*, is a member of the family Enterobacteriaceae, this therefore means that the *caf* operon can be expressed and F1 assembled efficiently in *Salmonella*. In addition, a study reported that a single dose of live attenuated *Y. pseudotuberculosis*, producing F1 from *Y. pestis*, gives protection against pneumonic plague (Derbise et al., 2012). The same protection was observed in *Salmonella* producing F1 (Oyston et al., 1995).

As plague is still endemic in some parts of the world and a potential bioweapon, there is continued interest in developing next-generation vaccines. The development of new vaccines should consider protection from all forms of plague, no adverse side effects and induction of long-term immunity from a minimum number of doses. Furthermore, there is a lot of interest in finding alternate antigens that could be used in combination with F1 and/or V Ag preparation, as well as in attenuated strains of *Y. pestis* and *Y. pseudotuberculosis*.

1.3 F1 capsular antigen

The F1 capsular antigen is one of the key components of plague vaccines (see above). It is produced at higher temperatures of 35-37°C and covers the bacterial surface making the

bacteria more resistant to phagocytosis. However, the F1 capsule is not expressed at temperatures of 26°C and below, for example in the flea gut (MacIntyre, 2004).

The F1 capsule consists of high molecular weight polymers built from a single protein subunit, Caf1 (15.6 kDa) which is encoded by the *caf* operon. The F1 polymer can easily extend to twice the diameter of the bacterial cell (MacIntyre et al., 2004). Assembly of Caf1 requires additional genes of the *caf* operon, which is located on the pFra/pMT1 plasmid. The first *caf* operon was cloned and sequenced from *Y. pestis* pEV76 (Galyov et al., 1990; Perry and Fetherston, 1997). The *caf* operon encodes four genes: *cafIR* (Zavialov et al., 2002), *cafI* (Galyov et al., 1990), *cafIA* (Perry and Fetherston, 1997) and *cafIM* (Brubaker, 1972). Early genome sequences indicated identity of each of these genes among different strains of *Y. pestis*. With the dramatic increase in genome sequencing capabilities it remains to be seen if there are any exceptions to this. Caf1R is synthesised as a 36.83kDa regulator and shares homology with the AraC family of regulators. The *cafI* gene encodes the structural subunit of the F1 polymer. Caf1 is synthesised as a 17.6 kDa precursor and is processed to a 15.56 kDa mature subunit during secretion across the inner membrane. This is not only an absolutely conserved surface protein, but Caf1 is also unique to *Y. pestis*. So far, no close homologues of Caf1 have been identified in the database. This is the reason that F1 and antibodies against F1 have proven to be so useful in diagnosis of plague and identification of *Y. pestis*. Both *cafIA* and *cafIM* are assembly related genes where Caf1M, originally assigned as Caf1 mediator, is synthesised as a 28.76 kDa precursor of a 26.3 kDa periplasmic chaperone. The *cafIA* gene encodes Caf1A, which is synthesised as a 93.0 kDa precursor of a 90.52 kDa outer membrane usher (MacIntyre et al., 2004). The Caf operon belongs to a well characterised chaperone-usher assembly pathway that will be discussed later.

The Caf1 subunits (15.56 kDa) assembly into fibres that collapse on the surface creating an antiphagocytic capsule-like structure that covers the cell surface. The

structure of the F1 fibre was derived from the crystal structure of a ternary Caf1M:Caf1:Caf1 complex (Zavialov et al., 2003). The thin linear F1 fibre with diameter of 20Å was assembled by Caf1 subunits (Figure 1.1) (Zavialov et al., 2003). The F1 polymer is remarkably stable, disassociates into individual oligomers and monomers only at high temperatures above 87.5°C (Zavialov et al., 2005). In 0.5% SDS, the fibre begins to dissociate at 75°C. The periplasmic chaperone Caf1M forms a donor strand exchange (DSC) complementation with the unfolded Caf1SC subunit (Zavialov et al., 2003). This drives folding of the Caf1 subunit in the periplasm, but this interaction is very different from the subunit:subunit interaction (see below) dissociating at a temperature of 55°C (Zavialov et al., 2005). In the absence of Caf1M interaction, unfolded Caf1 subunit is degraded in the periplasm.

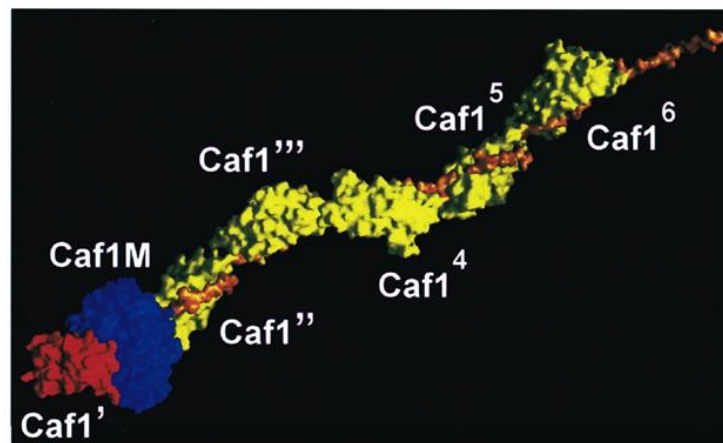


Figure 1.1: Architecture of F1 fibre. Blue indicates the chaperone that binds to incompletely-folded subunit (Red). Yellow indicates folded subunits that form a fibre and are exported to the cell surface. Orange indicates the terminal extension of Caf1 that helps to stabilise F1 polymer by donor strand exchange mechanism (Image taken from Zavialov et al., 2003).

1.4 Phylogeny of chaperone-usher (CU) families

CU systems are widespread, particularly among the *Enterobacteriaceae*. It is not uncommon for a single strain to possess different operons encoding CU assembled pili or fibres. While there is no evidence thus far of a role of F1 in adhesion, many CU pili are important virulence factors, due to their key roles in adhering to specific host cells (Zavialov et al., 2007; Wright et al., 2007). The chaperone-usher families are organised into gene clusters that encode usher, chaperone and major subunit proteins with variable addition of regulator and/or adhesin or other minor assembly subunits (Nuccio and Baumler, 2007). The major subunits assemble into thin flexible fimbriae that consist of two subunits per helical turn or thick rigid fimbriae that consist of 3 to 3.4 subunits per helical turn. Minor subunits generally assemble into the tip of the fimbriae and act as tip adhesin, tip fibrillum or/and adaptor that join the structure (Baga et al., 1987; Bullitt and Makowski, 1995; Jacob-Dubuisson et al., 1993; Kuehn et al., 1992). The usher proteins shared conserved domain of PFAM00577 and COG388 (Nuccio and Baumler, 2007). Despite the fact that the usher proteins share homology with several proteins from other systems, the chaperone/subunit complexes are very specific to their own usher (Russell and Orndorff, 1992; Saulino et al., 1998).

Of the many non-flagellar surface appendages, including curli, type IV pili, type III secretion pili and type IV secretion pili (Figure 1.2), chaperone-usher (CU) pili are probably the most extensively studied. One mode of classification was based on chaperone studies (Zav'yalov et al., 1995; Hung et al., 1996). The chaperone-usher family were classified into two main structurally and functionally distinct groups, FGL (long F1-G1 loop in the subunit binding domain) and FGS (short F1-G1 loop). These 2 groups corresponded to the assembly of thin fimbriae (fimbriae polyadhesins) and rod-like pili

(adhesive pili), respectively. Table 1.2 below shows a summary of the functional and structural classification of these fimbriae.

Table 1.2: Function-structure classification of chaperone-usher fimbrial adhesins.

Mono-adhesive binding refers to adhesive pili with only one specialised adhesive domain (adhesin). Polyadhesive binding refers to usually thin flexible fimbriae that possess one or two independent binding sites specific to its host cell receptor.

Function	Morphology	Chaperone	No. of subunits, assembled by chaperone	Adhesin subunit	Invasin or invasin-like subunit	References
Mono-adhesive binding	Thick rigid pili with a diameter of 7–8 nm	FGS: PapD, PrsD, FimC, SfaE, FocC, HifB, HafB, FimB, MrpD, PmfC, LpfB, PefD, AftB, FasB	5–7	One domain (adhesin), on the tip	Non-detected	Bullitt & Makowski (1995), Thanassi <i>et al.</i> (1998), Soto & Hultgren (1999), Knight <i>et al.</i> (2000), Sauer (2000, 2004)
Mono-adhesive binding	Thin flexible pili with a poorly defined diameter (2–4 nm)	FGS: F17D, MrkB, RalE	F17-2 Mrk-3 Ral-4	One domain (adhesin), on the tip	Non-detected	Lintermans (1988, 1991), Gerlach <i>et al.</i> (1989), Bullitt & Makowski (1995), Adams <i>et al.</i> (1997), Sebgathi <i>et al.</i> (1998), Soto & Hultgren (1999)
Polyadhesive binding	Thin flexible pili with a poorly defined diameter (2–4 nm)	FGS: FaeE, FanE	5	The main structural subunit: FaeG, FanG	Non-detected	Bakker <i>et al.</i> (1991), Soto & Hultgren (1999), van den Broeck (2000)
Polyadhesive binding	Atypical fimbrial structures	FGL, SafD-like family: CssC	2	The main structural subunit: CssA	The main structural subunit: CssB	(Di Yu <i>et al.</i> , 2012), (Remaut <i>et al.</i> , 2008)

Polyadhesive binding	Unusually long, thin (2 nm diameter) SEF18 fimbriae were evident on <i>Salmonella enteritidis</i> and <i>Escherichia coli</i> by electron microscopy	FGL, SafD-like family: SefB	2	The main structural subunit: SefA	The main structural subunit of alternative fimbriae: SefD (SEF18 fimbriae)	Clouthier (1993, 1994)
Polyadhesive binding	Atypical fimbrial structures: amorphous or capsule-like at low resolution	FGL, SafD-like family: SafB	2	The main structural subunit: SafA	One subunit on the tip: SafD	Folkesson <i>et al.</i> (1999), McClelland (2001, 2004), Deng <i>et al.</i> (2003), Chiu <i>et al.</i> (2005)
Polyadhesive binding	Thin flexible fibers (2 nm diameter) were observed for Dr adhesin by electron microscopy	FGL, Caf1M-like family: Afa-8B, AafD, Agg-3D, AggD, DafaB, Afa-3B, DraB, DaaB	1	The main structural subunit: Afa-8E, AafA, Agg-3A, AggA, DafaE, Afa-3E, DraE, DaaE	One subunit might be displayed on the tip after chaperone/usher independent secretion	Bilge <i>et al.</i> (1989), Garcia <i>et al.</i> (1994), Savarino <i>et al.</i> (1994), Elias <i>et al.</i> (1999), Lalioui & Le Bouguénec (2001), Bernier <i>et al.</i> (2002), Keller <i>et al.</i> (2002), Van Loy (2002), Le Bouguénec (2005), Piątek (2005), Servin <i>et al.</i> (2005), Zalewska <i>et al.</i> (2005), Jedrzejczak <i>et al.</i> (2006), Cota (2004, 2006)
Polyadhesive binding	From thin flexible fibers (2 nm diameter) observed for Psa fimbriae to amorphous or capsule-like morphology for F1 antigen (by electron microscopy)	FGL, Caf1M-like family: Caf1M, PsaB, MyfB, NfaE, CS3-E	1	The main structural subunit: Caf1, PsaA, MyfA, NfaA, CS3	Non-detected	Ahrens <i>et al.</i> (1993), Jalajakumare <i>et al.</i> (1989), (Galyov <i>et al.</i> , 1990), (Yu <i>et al.</i> , 2009), Iriarte <i>et al.</i> (1993), Lindler & Tall (1993), Yang <i>et al.</i> (1996), Zav'yalov (1996), Payne <i>et al.</i> (1998), Makoveichuk <i>et al.</i> (2003), Servin <i>et al.</i> (2005)

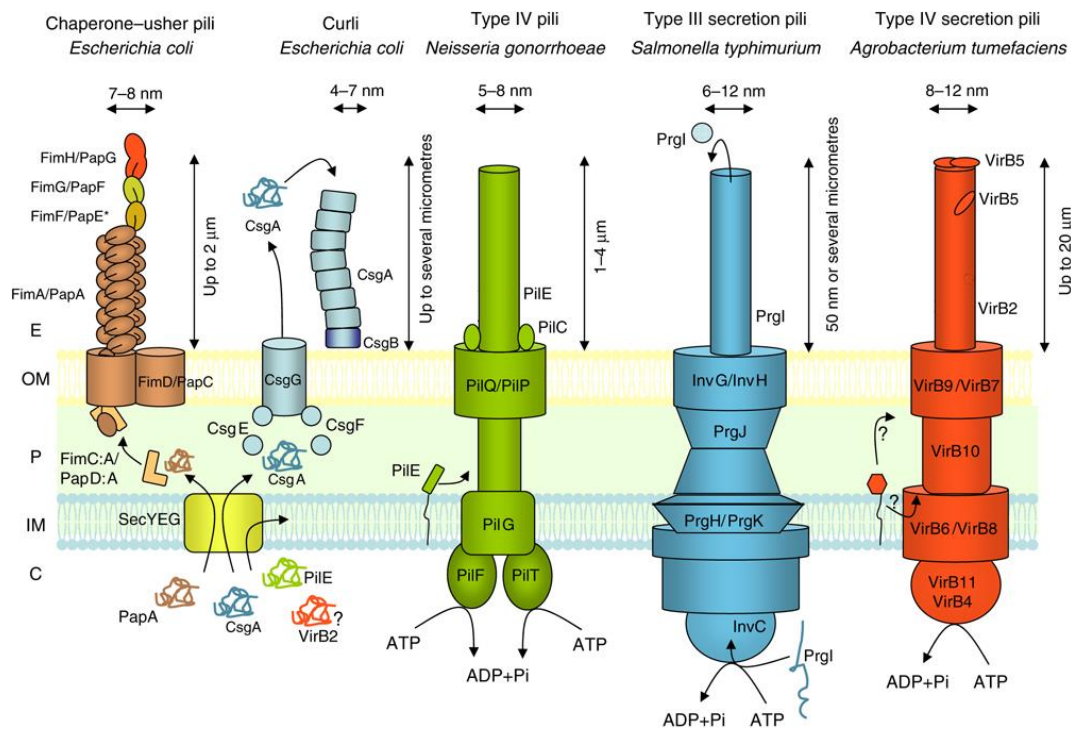


Figure 1.2: Assembly of CU pili and curli is driven by simple systems, involving specific periplasmic chaperones and outer membrane ushers. On the other hand, assembly of type IV pili, type III secretion pili and type IV secretion pili is driven by large multisubunit complexes crossing the whole bacterial cell envelope. (Taken from Waksman G., 2008)

As the number of sequenced CU systems greatly increased, Nuccio et al. (2007) classified CU systems according to the fimbrial usher protein (FUP) phylogeny. Ushers were classified into six major clades; α -fimbriae, β -fimbriae, γ -fimbriae (subdivided into γ_1 , γ_2 , γ_3), κ -fimbriae, π -fimbriae and σ -fimbriae (Figure 1.3). From this study, it became clear that the FGL systems form a small monophyletic group (γ_3 -fimbrial clade), whereas FGS systems fall into a much larger and diverse number of clades. The major FUP clade is σ -fimbriae where the members of this clade have limited or no sequence homology to the members of alternate fimbriae (α -fimbriae) or to classical CU family (β -, γ -, κ -, and π -fimbriae).

Type 1 pili are grouped together in γ_1 -fimbriae where it consists of the CU system from *Citrobacter freundii* (FimD), *E. coli* (AufC, FImD, Fm1C, FocD, SfaF,

SfmD, Ycbs and Ydet), *Proteus mirabilis* (AtfC) and *S. enterica* serotype Typhimurium (BcfC and FimD). Figure 1.3(B) shows that, with the exception of *P. mirabilis atf* operons of the γ 1-fimbriae clade this clade all encode usher, chaperone, more than one subunit and a typical tip adhesin. Most type 1 pili have been well studied and in particular the *E. coli fim* operon. It encodes a rigid fimbrial shaft of FimA major subunit and ends with the tip adhesin FimH, connected by FimF and FimG minor subunits (Figure 1.2) (Hahn et al., 2002; Jones et al., 1995; Saulino et al., 2000). *E. coli* type I pili bind to mannosylated glycoproteins in the bladder and is a key virulence determinant in cystitis (Sokurenko et al., 1994; Mulvey et al., 1998).

The γ 3-fimbrial clade, to which Caf1A usher belongs, also contains the *E. coli* ushers AfaC-3, AfaC-7, AfaC-8, AggC, Agg3C, CS3-2, CssD, DraC, and HdaC, PSPPH_A0064 of *Pseudomonas syringae*, SefC of *S. enterica* serotype Enteritidis, SafC of *S. enterica* serotype Typhimurium, PsaC of *Y. pestis* and MyfC of *Y. enterocolitica*. In general, γ 3-fimbriae do not contain specialized subunits with the 2-domain structure of tip adhesins. They are composed of only one major subunit and some are accompanied by non-polymerising subunit (Afa/Dra of *E. coli*) to cut their distal end. It assembles into thin flexible fibres of ~2-3 nm thickness and curl up on the cell surface to form capsule (Fronzes et al., 2008; Nuccio and Baumler, 2007). Caf1A, responsible for formation of the Caf1 Subunit into F1 fibres, is the best studied representative of the γ 3-fimbrial clade. The widely studied P and type I pili cluster in the π and γ 1- fimbrial clades, respectively. Pap and Fim chaperone usher systems from *E. coli* have been used as models to study FGS family and Caf chaperone usher system has been used as models of the γ 3/FGL system (Russell et al., 1995; Zavialov et al., 2007).

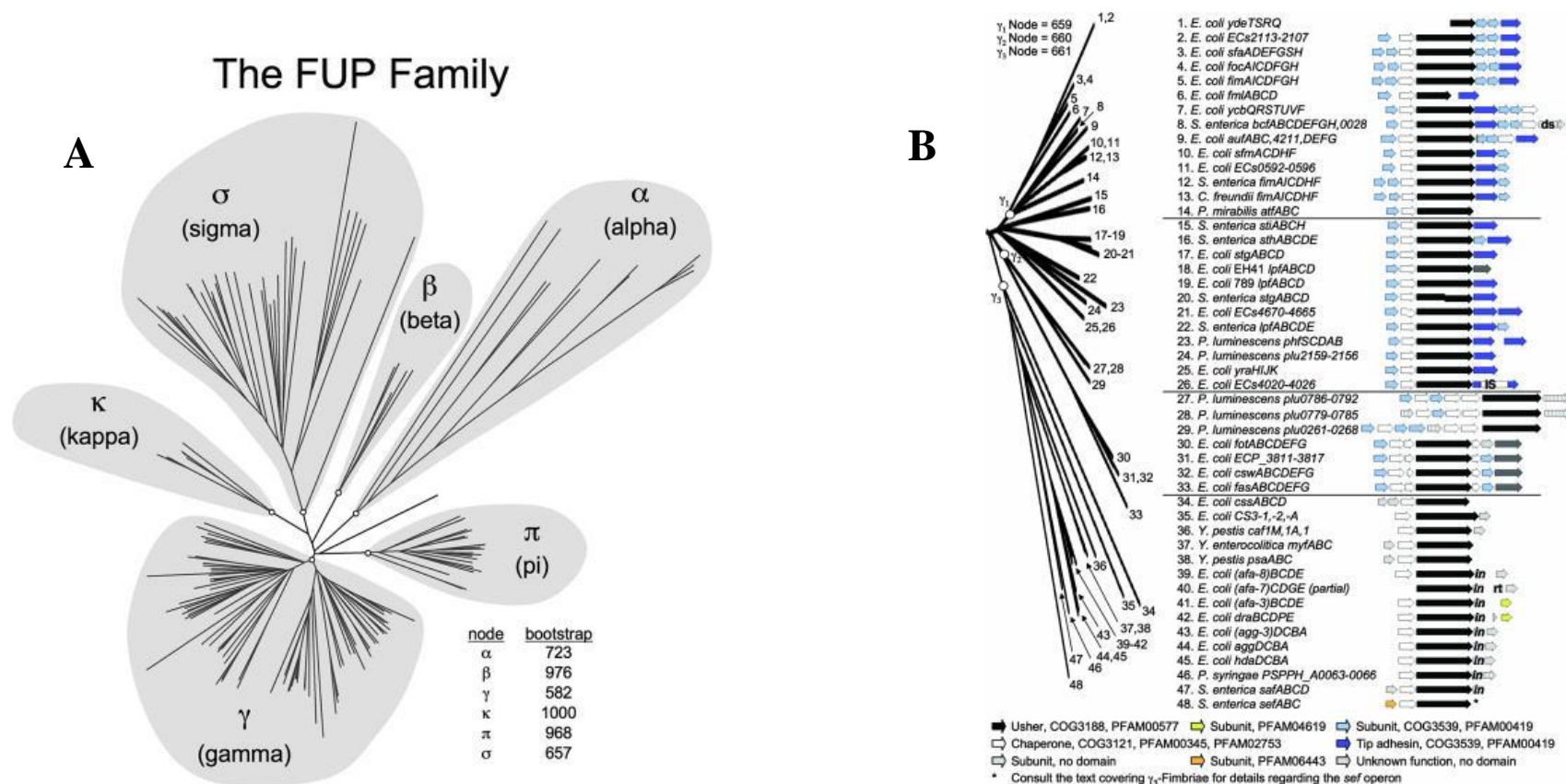


Figure 1.3: Phylogenetic tree of the fimbrial usher family.

(A) Phylogenetic tree generated from usher amino acid sequences. The ushers grouped into six fimbrial clades called α , β , γ , κ , π and σ . (B) Gamma fimbrial clade are divided into three subgroup; γ_1 , γ_2 and γ_3 . *Y. pestis* Caf1A usher lies in γ_3 group where it neighbored with myfABC (*Y. enterocolitica*) and psaABC (*Y. pestis*).

1.5 Assembly of F1 capsular antigen via chaperone-usher pathway

The assembly of F1 starts with the translocation of the Caf1 precursor from the cytoplasm to the periplasm via the general Sec pathway (Figure 1.4). The periplasmic chaperone binds to the subunit, aids folding and stabilises a folding intermediate by capping the assembly surface. This in turn prevents proteolytic degradation of unfolded subunit (Chapman et al., 1999; MacIntyre et al., 2001).

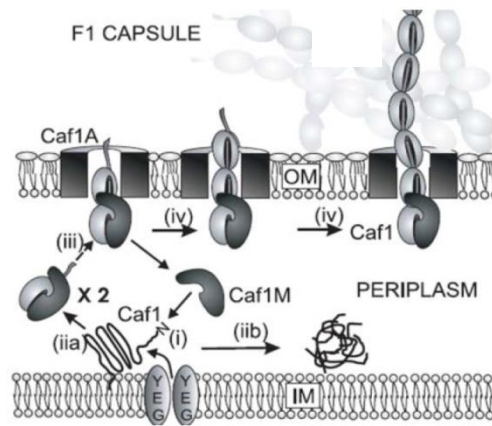


Figure 1.4: Summary of Caf CU pathway.

(A) Assembly of F1 polymer mediated by the chaperone (Caf1M) and usher (Caf1A) which provides a pore for the polymer to be transported outside the cell. Major steps: (i) subunit enters periplasm (iia) Chaperone capture/fold of newly synthesised pilin subunit in the periplasmic space, forms subunit:chaperone complex with incomplete subunit fold (iib) subunit that does not bind by chaperone will be degraded. (iii) Chaperone:subunit complex binds to the usher and opening the pore and. (iv) Multiple cycles of DSE where the chaperone G1 strand is displaced by a subunit Gd strand to form a complete folded subunit module within the fiber and chaperone recycling. (v) elongation of fiber. (Image taken from MacIntyre, 2004).

1.5.1 Donor strand complementation (DSC) of subunit by chaperone

Caf1 is bound to Caf1M to form an incomplete immunoglobulin (Ig) like structure with only six β -strands (A-F) of Caf1 (Figure 1.5). This process is called donor strand complementation (DSC) (MacIntyre et al., 2001; Yu et al., 2012). Caf1 subunits in the F1 polymer form a complete Ig fold where the missing seventh Gd strand is donated by the N-terminus of the adjacent subunit. In the absence of the seventh β -strand, the subunit would have a deep hydrophobic cleft (Sauer et al., 2002; Knight, 2007; Zavialov et al., 2003). This explains the instability of the free subunit, where it will aggregate and rapidly degrade.

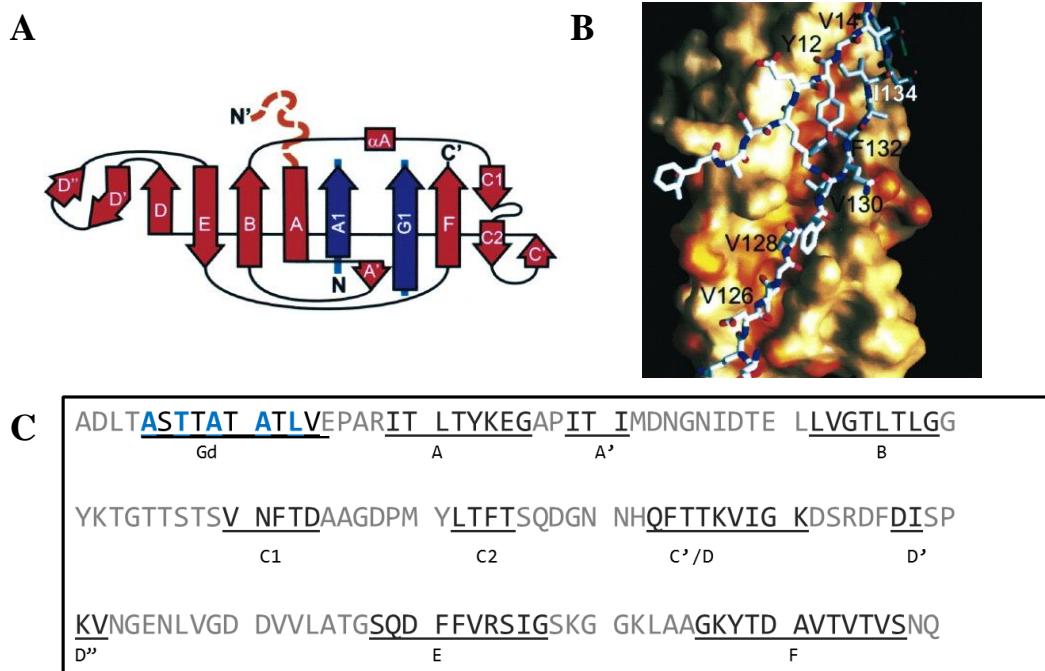


Figure 1.5: Topology Diagram and sequence of Caf1A.

(A) Topology diagram of chaperone bound Caf1. Caf1 monomer consists of six β -strands (A-F) with N-terminal extension. A1 and G1 (blue) are β -strands donated by Caf1M that stabilise the subunit (red) by hydrogen bonding to edge strands and filling the hydrophobic cleft. (B) The chaperone A1 and G1 filled the cleft of Caf1 (DSC) (Image of A and B taken from Zavialov et al., 2003). (C) Sequence of Caf1 subunit, 149 residues. Underlined residues indicate β -strand secondary structure with strand assignment given below each strand. Shaded in orange is the N-terminal extension Gd donor strand. Key donor residues (Ala5, Thr7, Ala9, Ala11 and Leu13) coloured in blue. (Image modified from MacIntyre et al., 2004).

Caf1M consists of two Ig-like domains. In the first domain, the two edge β -strands, A1 and G1 play an important role in stabilising and preventing subunit aggregation and degradation. Five hydrophobic residues in the G1 strand (Val126, Val128, Val130, Phe132 and 134Ile) which are larger than the corresponding Gd residues (Ala5, Thr7, Ala9, Ala11 and Leu13), bind to the subunit F strand in a parallel rather than the classic anti-parallel orientation (Figure 1.5 (B)). At the same time, the A1 strand from the chaperone stabilises the complex via edge-strand bonding to the subunit A strand in an anti-parallel interaction. The combination of edge strand interactions and hydrophobic residue/ hydrophobic cleft interactions results in a closed “super-barrel” and a Caf1:Caf1M intermediate that stabilises Caf1 subunit in an incompletely folded conformation (MacIntyre et al., 2001; Zavialov et al., 2003; Di Yu et al., 2012).

1.5.2 Donor strand exchange (DSE) at periplasmic surface of Caf1A usher

Caf1A is a typical β -barrel transmembrane outer membrane protein. It contains four additional functional domains: the N-terminal domain (NTD), a plug domain (structurally similar to Caf1 subunit) (Huang et al., 2009; Remaut et al., 2008; Yu et al., 2009) and two C-terminal domains (CTD1 and CTD2) (Dubnovitsky et al., 2010; Ford et al., 2010; Phan et al., 2011). The NTD and middle domain/plug initiate the assembly of Caf1 subunit by allowing chaperone:subunit complex to bind to NTD and moving the plug out and opening the Caf1A barrel. At this step, how the pore opens is still unclear. Nine residues in Caf1A NTD (Tyr2a, Thr3a, Phe4a, Thr7a, Met8a, Leu9a, Leu17a, Val19a and Phe22a) interact with seven residues in Caf1M (Pro4m, Leu43m, Leu67m, Ile102m, Pro103m, Pro104m and Arg69m) via Van der Waals interactions. The seven residues from Caf1M provide a large hydrophobic usher binding surface

(UBS). Interaction between Caf1A with the UBS contributes to about 97% of the interaction, whereas, Caf1 contributes only about 3% to the USB pre-assemble complex (Di Yu et al., 2012). The mechanism by which the usher specifically recognises loaded chaperone has recently been explained (Figure 1.6). In free Caf1M, Pro104m is in a collapsed conformation. This results in a closed UBS, preventing Phe4a (a key usher residue) interact with Caf1M because the position is occupied by Pro104m. However, when Caf1M is in a complex with Caf1, there is an allosteric induced conformational change that rotates Pro103m and Pro104m and moves Pro104m forming an open Phe4 for binding. This mechanism allows only the loaded incoming Caf1M:Caf1 complex to bind to usher (Di Yu et al., 2012).

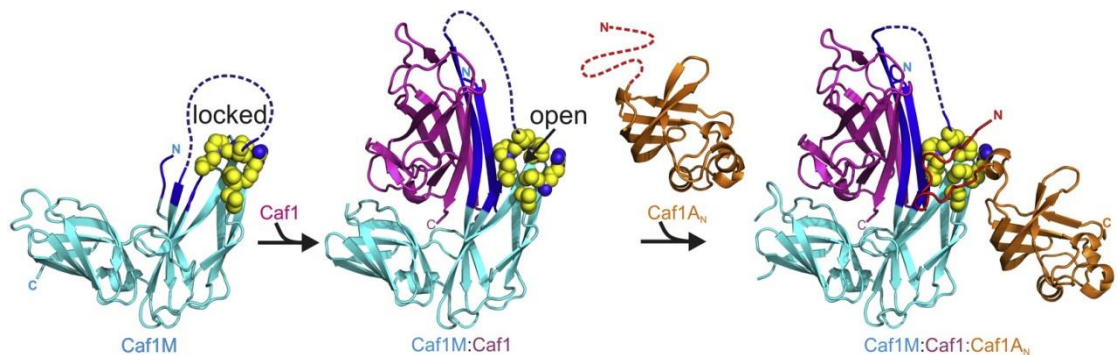


Figure 1.6: Caf1M:Caf1 complex allows proline lock (Pro104m) to open. In free Caf1M (cyan), UBS is closed by Pro104m (yellow). Caf1 (magenta) binds to Caf1M opening the UBS by moving Pro103m (yellow) and Pro104m. This allowing insertion of usher (orange) Phe4 to the open UBS, stabilising chaperone-usher interaction (Image taken from Di Yu et al., 2012).

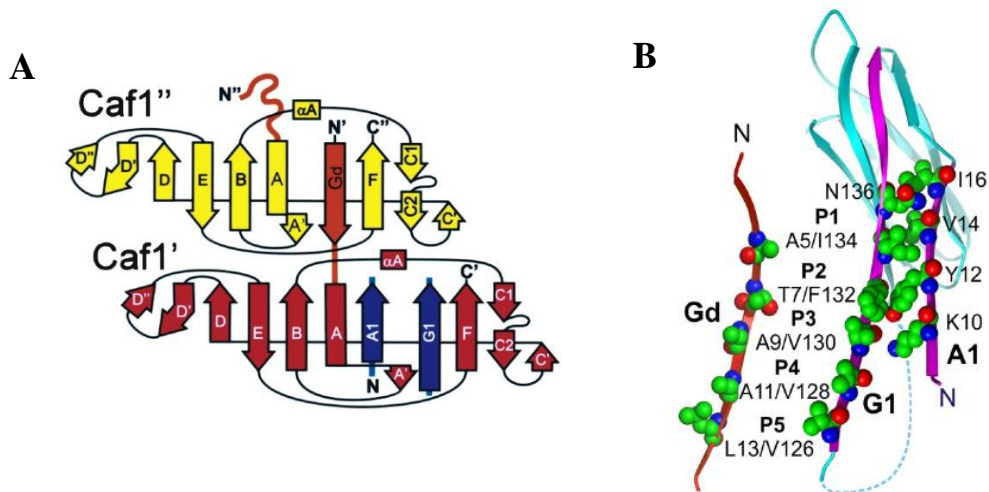


Figure 1.7: Topology diagram of Caf1M:Caf1:Caf1 interactions.

(A) The N-terminal extension of Caf1' subunit (orange) fills the cleft following displacement of the chaperone G1 strand by DSE. This completes the Ig fold of Caf1''. (Zavialov et al., 2003). (B) Ribbon diagrams of Gd and G1 donor strands. Chaperone G1 and A1 strands in violet and subunit Gd strand in red. The G1 and Gd donor residues are shown in ball-and-stick labelled according to subunit acceptor cleft (P1-P5) (Image taken from MacIntyre, 2004).

Donor strand exchange (DSE) takes place when Caf1 N-terminal extension (Gd) from an incoming chaperone:subunit complex (Caf1M:Caf1'') replaces the chaperone A1 and G1 strand from the chaperone:subunit complex (Figure 1.7(A) (Sauer et al., 2002; Zavialov et al., 2002). As shown in Figure 1.7(B), there are five pockets in the Caf1 hydrophobic cleft. The DSE is initiated at P5 pocket of the subunit where the P5 pocket, filled with Val126m is displaced by the corresponding Gd residue from Caf1'' via zipper-like mechanism (Remaut et al., 2006; Zavialov et al., 2003). Binding of the Gd strand to the cleft forms a complete seven-stranded Ig fold. The Gd complemented subunit undergoes a conformational change during DSE, changing from a high-energy, incomplete folded chaperone bound form to a compact, highly stable, fibre inserted state (Zavialov et al., 2003; Zavialov et al., 2005). The translocation of the fiber to the cell surface is probably driven by the folding energy released (Zavialov et al., 2003; Yu

et al., 2012). The polymer keeps growing via additional DSE interactions of chaperone-subunit complexes, leading to the localisation of the growing polymer through the usher protein to the cell surface. A fibre of F1 polymer can be built from more than 100 subunits. The F1 fibres on the cell surface collapse into a capsule-like structure (Yu et al., 2012; Zavialov et al., 2002).

1.6 Assembly at the *E. coli* ushers PapC and FimD

The most extensively studied chaperone-usher pathways are the Pap and Fim systems in uropathogenic *E. coli* (UPEC) which assemble type P and type I pili, respectively (Busch and Waksman, 2012). Type I pili are encoded by the *fim* gene cluster (*fimA-fimI*) where *fimH* encodes an adhesin responsible for targeting the bacteria to epithelial bladder and kidney cells (Korea et al., 2011). Type P pili are encoded by the gene cluster (*papA-papK*) where *papG* encodes the tip adhesin that promotes binding of the bacteria to kidney epithelial cells and erythrocytes.

The assembly of pili via the Fim and Pap systems were first described for the FimC-FimH (Choudhury et al., 1999) and PapD-PapK (Sauer et al., 1999). The basic principle of assembly of Pap and Fim systems are similar to that of the Caf system except Pap and Fim systems assemble more than one type of subunit to form structured pili. In the Pap system, the assembly starts with adhesin protein PapG, followed by the adaptor subunit PapF, 10 copies of flexible tip fibrillum PapE, a single copy of another adaptor subunit PapK, a rigid rod consisting of a thousand units of PapA followed by the termination subunit PapH (Verger et al., 2006). On the other hand, in the Fim system, only 4 subunits are required to form a type I pili. It starts with the assembly of adhesin subunit FimH, one copy each of FimG and FimH to form the flexible tip fibrillum. This

is then followed by approximately 1000 copies of FimA. No termination subunit has yet been characterised in the Fim system (Hahn et al., 2002; Le Trong et al., 2010).

The first structure solved for any usher membrane domain was that of the barrel domain of PapC in 2008 (Remaut et al., 2008). This 130-640 amino acid fragment crystallised as a 24-stranded β -barrel (residues 146-635), 45 Å height, 65-45 Å outer dimension and 45-25 Å outer dimension (PDB:2VQI, 3.2 Å). In this structure, between transmembrane β 6 and β 7 forms a 6 stranded β sandwich fold (the plug or middle domain) located inside the barrel. As shown in Figure 1.8 (C). The β 5-6 hairpin residues (Ile231 and Phe236) bend in towards the barrel and interact with residues on the barrel wall (Trp239, Phe357 and Val377) and α -helix (Met451) through hydrophobic interactions. There are also two electrostatic interactions that connect the β 5-6 hairpin with the barrel wall (Glu361-Lys339-Asp234-Arg303) and β 5-6 hairpin with α -helix (Arg237-Glu467-Arg305-Asp32) Figure 1.8 (C). These electrostatic interactions help the plug reside inside the PapC barrel (Remaut et al., 2008).

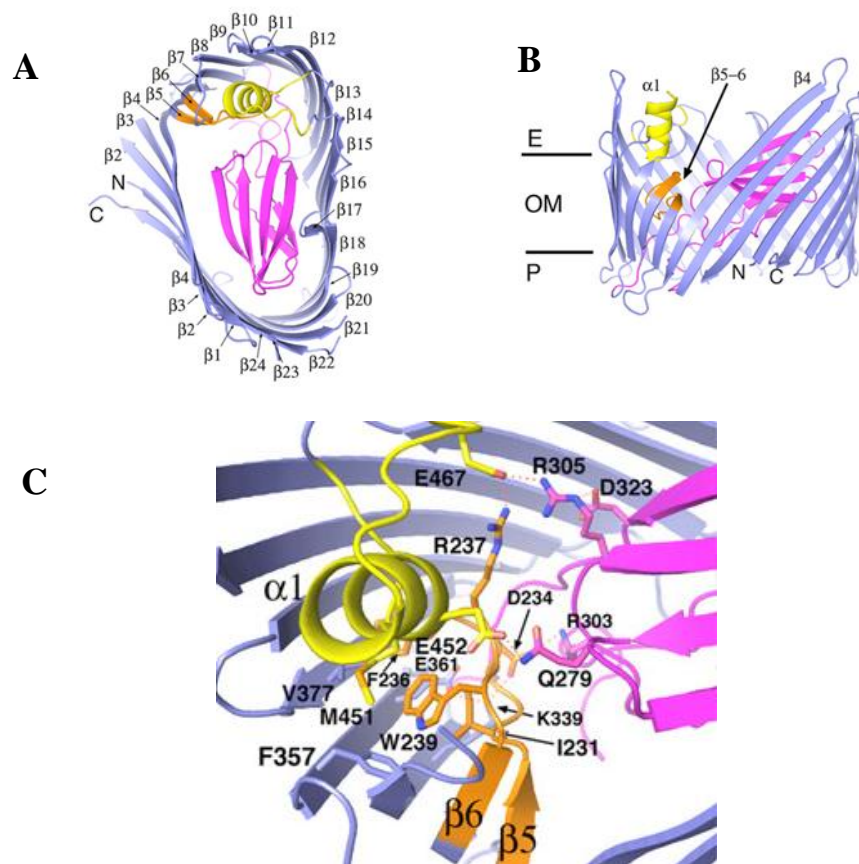


Figure 1.8: Crystal structure (3.2Å resolution) of PapC, PDB: 2VQI.

(A) View from top of PapC usher showing all 24 β -strands. N and C indicate N and C terminus. (B) View from side of PapC usher. The β -barrel coloured in blue, plug domain in magenta, β 5-6 hairpin in orange and α -helix in yellow. (C) Hydrophobic and electrostatic interactions between β -barrel, β 5-6 hairpin and α -helix. (Image taken from Remaut et al. 2008).

For the subunit fibre to go through the PapC barrel, the middle domain must move out of the barrel into the periplasm. It has been proposed that the binding of the chaperone: adhesion complex, induced activation of the usher (Saulino et al., 1998). Opening of the barrel would create a $37 \times 25\text{\AA}$ translocation channel or a channel of $45 \times 25\text{\AA}$ when α -helix or β 5-6 hairpin are displaced from the barrel. It was suggested that the required conformational rotation of the plug is driven by energy coming from binding of the chaperone-adhesin complex (Remaut et al., 2008).

In 2011, the crystal structures of FimD bound to the FimC:FimH complex and the unbound usher were solved (Phan et al., 2011). FimD is the only usher for which a high-resolution structure has been solved with a subunit inserted into the pore. This study highlighted that the inserted FimC:FimH complex was bound to C-terminal domain 1 (residues 666-750) and to C-terminal domain 2 (residues 751-834). The ternary complex also shows that the plug domain (residues 241-324) was not pushed out of the pore lumen as suggested earlier (Remaut et al., 2008) but resides on the periplasm face of the usher in the loaded translocon (Phan et al., 2011; Yu et al., 2009). The correct ordering of the subunits has been attributed in part to the high, but differing affinity of subunit:chaperone complex for the PapC N-terminal domain (Sauer et al., 1999; Dodson et al., 1993).

There are now three solved structures of FimD, which are FimD membrane barrel (in closed state) without any periplasmic domain (PDB:3OHN, 3.01 Å) (Phan et al., 2011), FimD bound to FimC:FimH complex accommodating FimH subunit in the barrel (PDB:3RFZ, 2.8 Å) (Phan et al., 2011) and FimD bound to FimC:FimF:FimG:FimH complex accommodating FimH subunit in the barrel (PDB:4J3O, 3.8 Å) (Geibel et al., 2013). In 2011, Phan et al. reported that activation of the FimD usher on complexing FimC:FimH drives a dramatic conformational changes to the FimD barrel from diameter of 52 Å by 28 Å and near oval-shaped pore to 44 Å by 36 Å and a more circular pore (see Figure 1.9 (A)). This correlates with a change of closed to open state (removal of the plug), and insertion of the adhesin into the usher pore. Apparent flexibility of the β -barrel had not previously been seen in this or any other β -barrel outer membrane protein (Phan et al., 2011). Based on the paper published in 2013 by Geibel et al., upon translocation of type 1 pili, counterclockwise rotation of

approximately 110° to 120° and an upward motion of approximately 50 Å of FimH was observed (see Figure 1.9 (B)). This hypothesized that the rotation prevents FimH from sliding back into the periplasm (Geibel et al., 2013).

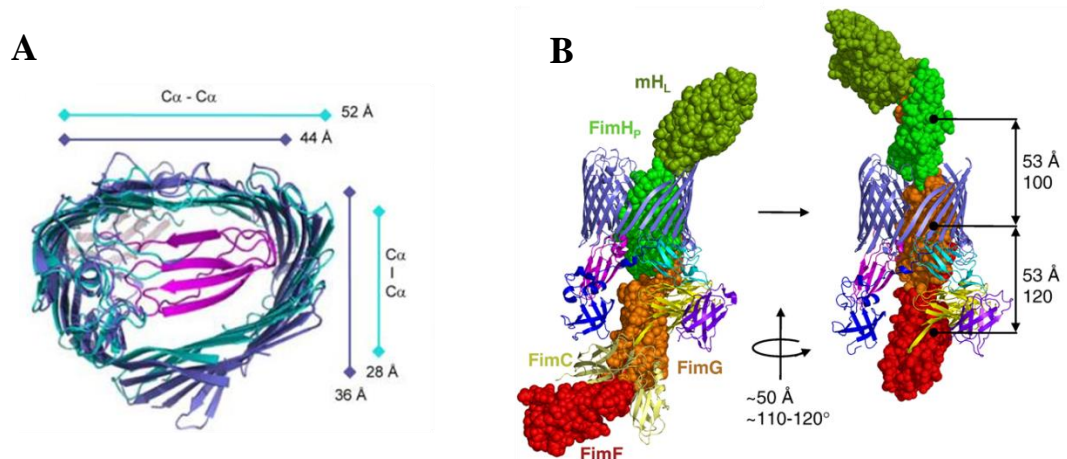


Figure 1.9: Closed and open structures of FimH. (A) Superimposed and comparison of FimD close and opened structure. (B) The rotation of the pili when FimD accommodates FimH and FimG (Image taken from Phan et al., 2011 and Geibel et al., 2013).

Available structures of the Caf system include, wildtype (PDB:1Z9S, 2.2 Å) and mutant (PDB:1P5U, 1.99 Å) of Caf1M:Caf1:Caf1 complexes, (Zavialov et al., 2005), the Caf1A C-terminal domain (PDB:2XET, 1.6 Å) (Dubnovitsky et al., 2010), recombinant Caf1A plug or soluble middle domain (PDB:3FCG, 2.85 Å) (Yu et al., 2009) and subunit-free Caf1M (PDB:4AY0, 1.52 Å) as well as the Caf1A N-terminal domain alone (PDB:4B0E, 2.0 Å) and in complex with Caf1M:Caf1 (PDB: 4B0M, 1.8 Å) (Di Yu et al., 2012). The only remaining unsolved structure in Caf system is the Caf1A usher barrel.

Deletion of the Caf1A N-terminus (Caf1A_N) showed a dramatic decrease in the production of F1 on the cell surface. The deletion of Caf1A Asn13-Gln28 and residues

Thr12-Pro99 decreased and abolished surface F1 production respectively thus suggesting that Caf1A_N is crucial in assembly or may disrupt the structure of the domain (Pudney, 2006). The PapC_N was suggested to have a recognition site for chaperone:subunit complexes. Ng et al (2004) also showed the substitution of Phe3 to Ala and deletion of residues 2-11 in the PapC_N abolished pilus assembly (Ng et al., 2004). It was also shown that the deletion of the FimD N-terminal tail (residues 1-24) completely abolished FimD function, suggested that the N-terminal tail contributes to recognition of chaperone subunit complex in the Fim CU pathway (Nishiyama et al., 2005).

Site-directed mutagenesis of Caf1A_N residues involved in interaction with Caf1M, Phe5Ala and Phe23Ala dramatically decreased levels of F1 on the cell surface. Caf1A_N Met9Ala also reduced surface F1 but not as dramatically as Caf1A_N-Phe5 and -Phe23. This suggest that Caf1A_N-Met9 is less important in binding with chaperone:subunit complex. These two important residues, namely Phe5 and Phe23 correspond to Phe4 and Phe22 in FimD_N and Phe3 and Phe21 in PapC_N. Mutation of FimD_N-Phe4Ala abolished binding of FimC-FimH explaining the loss of pili assembly (Nishiyama et al., 2005). Mutation of PapC_N-Phe3Ala also identified a critical requirement of this residues for binding Pap chaperone:subunit.. Mutation of PapC_N-Phe21Ala showed that it can assemble tips (PapG) in the absence of major pili PapA but cannot assemble pili. It was therefore proposed that Phe21 is involved in assembly of an ordered pilus (Li et al., 2010). A conserved region (disulphide loop) in FimD_N (residues 70-79) and PapC_N (residues 77-86) was deleted and shown to abolished catalytic activity of the usher. Mutations of FimD_N disulphide loop was able to bind chaperone:subunit complex, initiating assembly of pili but could only form a short pilus tip. However, mutations of

PapC_N disulphide loop was unable to form pili. These observations proposed that catalytic activity is critical to initiate pilus biogenesis in PapC and important for pilus extension in FimD (Henderson et al., 2011). The different functions of the catalytic activity in FimD and PapC may reflect the phylogenetic distance between these two ushers which are classified into γ and π FUP family, respectively. All of the evidence above suggests that the N-terminal domain is required for loaded chaperone to bind to usher and the middle domain

1.7 Aims and objectives

The aim of this project was to use a bioinformatics approach to further elucidate the Caf specific translocation process and to then apply knowledge of this to optimise export of additional short peptide sequences within the F1 scaffold. Strategic significance of this approach is related to the central importance of F1 in both diagnosis and prevention of plague (Perry and Fetherston, 1997). By using the available high-resolution crystal structures of the Caf subunit:chaperone complex (Zavialov et al., 2005) and the Fim translocon (Geibel et al., 2013; Phan et al., 2011) to model the Caf translocon, *in silico* analysis could be used to identify predicted conserved features and to guide the design of modified surface F1 polymer. This bioinformatics approach has then been readily complemented with established laboratory methods to monitor the efficiency of assembly of F1 carrying inserted peptides. During the search for additional γ 3 FUP family members, an usher sequence closely related to Caf1A usher was identified in a plasmid sequence from a commensal strain of *E. coli*. Because of the current interest in evolution of *Y. pestis* to a flea-borne route of transmission (Hinnebusch, 2005) and the use of F1-based serotyping in plague (Simon et al., 2013; Schofield et al., 2012) detection, the relationship between this *E. coli* CU system and *Y.*

pestis Caf system was also investigated. The main specific objectives of the study were as follows:

1. To characterize production of recombinant F1 from the low copy number plasmid pACYCF1
2. To extend the $\gamma 3$ family of fimbrial usher proteins and in particular identify any closely related to Caf1A
3. To model the Caf translocon using the available structures of the Fim translocon and the Caf subunit:chaperone complex
4. To use the model of the Caf translocon and multiple sequence alignments of related ushers to identify conserved residues that might influence translocation of F1 through the usher and to test this experimentally by mutagenesis.
5. To monitor the flexibility and efficiency of four different loops within F1 to act as replacement sites for short peptides.
6. Using the modelled translocon structure, identify and design repair of export 'blocks' of modified F1 polymer
7. Bioinformatic analysis of the *E. coli* SE11 chaperone usher cluster, the most closely related usher to Caf1A.
8. Demonstration of functionality of SE11 CU locus from pSE11-2 by recombinant production and purification of surface *E. coli* SE11 subunit(s) polymer.

CHAPTER 2 MATERIAL AND METHODS

2.1 Bacteria strain

Table 2.1. below indicates details of *E. coli* strains used throughout this study. Initially, DH5 α competent cells were used for protein expression and sequencing of site-directed and deletion mutants. StellarTM Competent cells were used to perform InFusion cloning and sequencing of the infused plasmid. Both *E. coli* BW25113 and TOP10 were used for transformation and protein expression of loop mutations, as indicated. *E. coli* BL21(DE3) was used to express the recombinant gene from pACYCD1_MA_S123 construct.

Table 2.1 List of competent cells used in this study

Strain	Genotype	Reference
DH5α	<i>endA1 hsdR17(r_k- m+k+)supE44 thi - 1 recA1 gyrA (Nalr)relA1 Δ(lacIZY A- argF)U169 deoR (ϕ80dlacΔ(lacZ)M15)</i>	Sambrook et al., 1989
TOP10	<i>F- mcrA Δ (mrr-hsdRMS-mcrBC) ϕ80lacZΔM15 ΔlacX74 recA1 araD139Δ(ara-leu) 7697 galU galK rpsL (StrR) endA1 nupG λ-</i>	Invitrogen (Cat No. C4040)
BW25113	<i>F- Δ (araD-araB)567 Δ lacZ4787 (::rrnB-3), lacIp-4000(lacIQ) λ- rph-1 Δ (rhaD-rhaB)568 hsdR514</i>	Prof. Simon Andrew stock (SCE960)
StellarTM Competent Cells	<i>F-,endA1, supE44, thi-1, recA1, relA1, gyrA96, phoA, Φ80d_ lac ZΔM15, Δ(lacZYA- argF) U169, Δ (mrr- hsd RMS -mcrBC), ΔmcrA, λ-</i>	Clontech Takara, UK (Cat No. 636763)
BL21 (DE3)	<i>fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5</i>	New England Biolab (Cat. No. C2527I)

2.2 Media and culture conditions

All *E. coli* strains were grown in Luria-Bertani broth (LB) (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.09 M NaCl pH 7.0) and Luria-Bertani agar (LA) (LB + 1.5 % w/v agar). For expression of F1, single transformants or subcultures from -80°C

transformant stocks were inoculated into 10 ml LB supplemented with 0.6% glucose, and appropriate antibiotics (Chloramphenicol, 34 µg/ml or Ampicillin, 100 µg/ml) and were incubated overnight (20-22 hours) at 26°C with shaking at 225 rpm. For induction from pACYCF1, 1 ml of the overnight culture was centrifuged for 10 min at 13000 rpm. Pelleted cells were washed in 1 ml LB at 13000 rpm 10 minutes, resuspended in 10 ml LB with 10 µg/ml chloramphenicol and thermoinduced at 37°C for 4 hours. For induction from pFMA1, a transformant colony was inoculated in 10mL LB with 0.6% glucose and 100µg/ml ampicillin for ON at 26°C. The next day, the sample was washed and subcultured 1/10 in 10mL LB with 100µg/ml AMP for 1 hour at 37°C followed by induction of 0.5 mM IPTG and incubated for further 1.5 hours at 37°C. Stock cultures were prepared by adding 700 µl of the 26°C overnight cultures to 300 µl of 50% glycerol stock (final concentration 15% w/v glycerol) and stored at -80°C

2.3 Plasmids and mutants

Table 2.2 provides details of all plasmids and constructs used in the study. The *caf* operon was expressed from either pACYCF1, pFMA1 plasmid, or derivatives thereof. The pACYCF1 plasmid expresses a good level of Caf1 protein and a low level of Caf1M which mimics the native *caf* operon. In pFMA1, there is a high level of Caf1M and a slightly lower level of expression of Caf1 from this plasmid. Expression of Caf1 from pFMA1 is toxic to *E. coli* after 2-3 hours (Kersley et al., 2003).

Table 2.2: Plasmids used in this study

Plasmid	Description	Reading stock number	Reference
pACYCF1	pACYCDuet with the entire <i>caf</i> system, subcloned into <i>NcoI</i> and <i>SacI</i> restriction sites of pACYC, under control of <i>caf1R</i> . Thermo-induced for 4 hours at 37°C	R1122	Mary Leonard and Dharmender Kumar, unpublished
pFMA1 (R352)	P _{trc99a} with <i>caf1M</i> <i>caf1A</i> , <i>caf1</i> under control of <i>ptrc</i> promoter. IPTG induced, 1 hour without glucose than 0.5mM IPTG for 1.5 hours at 37°C	R352	(Chapman et al., 1999)
pACYCF1-A-F4A	Derivative of pACYCF1 with N-terminal domain mutation Phe4 to Ala	R1156	This study
pACYCF1-A-F4S	Derivative of pACYCF1 with N-terminal domain mutation Phe4 to Ser	R1157	This study
pACYCF1-A-Y240A	Derivative of pACYCF1 with middle domain mutation Tyr240 to Ala	R1158	This study
pACYCF1-A-ΔN(T8-Q131)	Derivative of pACYCF1 with N-terminal deletion from Thr8 to Gln131. Used as control	R1159	This study
pFM1-F1d11 clone 4	D33N34 of Caf1 replaced with GG	R665	Elham Moslehi, Unpublished
pFMA1-F1D12 clone 1	S76QDGNHH82 of Caf1 replaced to GGGG	R784	Elham Moslehi, Unpublished
pFM1-F1D13 clone A	D92SRDF96 of Caf1 replaced to GGGG	R669	Elham Moslehi, Unpublished
pFM1-F1D14 clone A)	K129GGK132 of Caf1 replaced with DGGG	R670	Elham Moslehi, Unpublished
pFM1-F1::1 clone 1	D92SRDF96 of Caf1 replaced to ELDKWA	R899	Elham Moslehi, Unpublished
pFM1-F1::2 clone 1	S76QDGNHH82 of Caf1 replaced with ELDKWA	R900	Elham Moslehi, Unpublished

pFM1-F1::3 clone 2	D33N34 of Caf1 replaced to ELDKWA	R901	Elham Moslehi, Unpublished
pFM1-F1::4 clone 2	K129GGK132 of Caf1 replaced to ELDKWA	R902	Elham Moslehi, Unpublished
pACYC-F1D11	Encodes Caf1 loop 1 mutation D33N34→GG InFused from R665 into pACYCF1	R1161	This study
pACYC-F1D12	Encodes Caf1 loop 4 mutation S76QDGNHH82→GGGG InFused from R784 into pACYCF1	R1162	This study
pACYC-F1D13	Encodes Caf1 loop 5 mutation D92SRDF96→GGGG InFused from R669 into pACYCF1	R1167	This study
pACYC-F1D14	Encodes Caf1 loop 7 mutation K129GGK132→DGGG InFused from R670 into pACYCF1	R1160	This study
pACYC-F1::1	Encodes Caf1 loop 1 mutation D92SRDF96→ELDKWA InFused from R899 into pACYCF1	R1163	This study
pACYC-F1::2	Encodes Caf1 loop 4 mutation S76QDGNHH82→ELDKWA. InFused from R900 into pACYCF1	R1164	This study
pACYC-F1::3	Encodes Caf1 loop 5 mutation D33N34→ELDKWA. InFused from R901 into pACYCF1	R1165	This study
pACYC-F1::4	Encodes Caf1 loop 7 mutation K129GGK132→ELDKWA InFused from R902 into pACYCF1	R1166	This study
pACYCF1_A_Y266A	Derivative of pACYCF1 with middle domain mutation Tyr266 to Ala	R1220	This study
pACYCF1_A_Y266S	Derivative of pACYCF1 with middle domain mutation Tyr266 to Ser	R1221	This study
pACYCF1_A_Q169A	Derivative of pACYCF1 with usher mutation Gln169 to Ala	R1200	This study
pACYCF1_A_Q169G	Derivative of pACYCF1 with usher mutation Gln169 to Gly	R1201	This study
pACYCF1_F1_N80G	Derivative of pACYCF1 with <i>caf1</i> mutation Asn80 to Gly	R1202	This study
pACYCF1_F1_N81G	Derivative of pACYCF1 with <i>caf1</i> mutation Asn81 to Gly	R1203	This study

pACYCF1::4_F1_N81G	Derivative of pACYCF1::4 with <i>caf1</i> mutation Asn81 to Gly	R1204	This study
pACYCF1::2_F1_ELDK WANP	Derivative of pACYCF1::2 with Asn and Pro insertions in Caf1 insert loop 4	R1205	This study
pACYCF1::1_A_Q169A	Derivative of pACYCF1::1 with usher mutation Gln169 to Ala	R1206	This study
pACYCF1::1_A_Q169G	Derivative of pACYCF1::1 with usher mutation Gln169 to Gly	R1207	This study
pACYCF1_F1_P16A	Derivative of pACYCF1 with <i>caf1</i> mutation Pro16 to Ala	R1208	This study
pACYCF1::3_F1_P16A	Derivative of pACYCF1::3 with <i>caf1</i> mutation Pro16 to Ala	R1211	This study
pACYCF1_F1_R18K	Derivative of pACYCF1 with <i>caf1</i> mutation Arg18 to Gln	R1209	This study
pACYCF1::3_F1_R18K	Derivative of pACYCF1::3 with <i>caf1</i> mutation Arg18 to Gln	R1212	This study
pACYCF1::2_A_ELDK WANH	Derivative of pACYCF1::2 with Asn and His insertions in Caf1 insert loop 4	R1214	This study
pACYCF1_A_S182A	Derivative of pACYCF1 with usher mutation Ser182 to Ala	R1215	This study
pACYCF1::1_A_S182A	Derivative of pACYCF1::1 with usher mutation Ser182 to Ala	R1216	This study
pACYCF1_A_Q167S	Derivative of pACYCF1 with usher mutation Gln167 to Ser	R1217	This study
pACYCF1_A_Q167A	Derivative of pACYCF1 with usher mutation Gln167 to Ala	R1218	This study
pACYCF1::4_F1_N80G	Derivative of pACYCF1::4 with <i>caf1</i> mutation Asn80 to Gly	R1219	This study
pACYCF1D12_GGGGN NH	Derivative of pACYCF1D12 with Asn, Asn and His insertions in Caf1 insert loop 4	R1222	This study
pACYCF1::2_ELDKWA NNH	Derivative of pACYCF1::2 with Asn, Asn and His insertions in Caf1 insert loop 4	R1223	This study
pACYCF1_plug_E298K	Derivative of pACYCF1 with usher mutation Glu298 to Lys	R1240	This study
pACYCF1_plug_E291K	Derivative of pACYCF1 with usher mutation Glu291 to Lys	R1242	This study
pACYCF1_plug_E291A	Derivative of pACYCF1 with usher mutation Glu291 to Ala	R1242	This study
pACYCF1_plug_R263A	Derivative of pACYCF1 with usher mutation Arg263 to Ala	R1243	This study
pACYCF1_hairpin_Y21 5A	Derivative of pACYCF1 with usher mutation Tyr215 to Ala	R1244	This study
pACYCF1_hairpin_D21 7A	Derivative of pACYCF1 with usher mutation Asp217 to Ala	R1245	This study

pACYCF1_plug_S289A	Derivative of pACYCF1 with usher mutation Ser289 to Ala	R1246	This study
pACYCF1_plug_R258A	Derivative of pACYCF1 with usher mutation Arg258 to Ala	R1247	This study
pACYCF1_barrel_K462A	Derivative of pACYCF1 with usher mutation Lys462 to Ala	R1248	This study
pACYCF1_barrel_N144A	Derivative of pACYCF1 with usher mutation Asn144 to Ala	R1249	This study
pACYCF1_barrel_N148A	Derivative of pACYCF1 with usher mutation Asn148 to Ala	R1250	This study
pACYCF1_barrel_N601A	Derivative of pACYCF1 with usher mutation Asn601 to Ala	R1251	This study
pACYCF1_barrel_Q150A	Derivative of pACYCF1 with usher mutation Gln150 to Ala	R1252	This study
pACYCF1_barrel_Q464A	Derivative of pACYCF1 with usher mutation Gln464 to Ala	R1253	This study
pACYCD1_MA_S123	Encodes <i>E. coli</i> SE11 whole operon cloned into pACYCDuet1 (without upstream and downstream). IPTG induce.	R1193	This study
pACYCD1_MA_S1	Derivatives of R1193 encodes <i>E. coli</i> SE11 chaperone, usher and subunit 1 by deletion of subunit 2 and 3 (without upstream and downstream). IPTG induce.	R1197	This study
pACYCD1_MA_S23	Derivatives of R1193 encodes <i>E. coli</i> SE11 chaperone, usher, subunit 2 and 3 by deletion of subunit 1 (without upstream and downstream). IPTG induce.	R1234	This study
pTRC99-Mh-S123	Encodes <i>E. coli</i> SE11 chaperone, subunit 1, 2 and 3 cloned into pTRC99 with 6 His added at chaperone C-terminus.	R1195	This study
pTRC99-Mh-S1	Derivatives of R1195 encodes <i>E. coli</i> SE11 chaperone (6 His), and subunit 1 by deletion of subunit 1 and 2	R1231	This study
pTRC99-Mh-S23	Derivatives of R1195 encodes <i>E. coli</i> SE11 chaperone (6 His), subunit 1 and subunit 2 by deletion of subunit 1.	R1232	This study
pTRC99-Mh-S3	Derivatives of R1195 encodes <i>E. coli</i> SE11 chaperone (6 His), subunit 3 by deletion of subunit 1 and subunit 2.	R1198	This study

pTRC99-M8h-S123	Derivatives of R1195 encodes <i>E. coli</i> SE11 chaperone, subunit 1, 2 and 3 with insertion of 2 His making 8 His at chaperone C-terminus.	R1196	This study
pACYCF1::1_DS_6His	Derivative of pACYCF1 with 6 His insertion at C-terminus of Caf1M	R1225	This study
pACYCF1::1_DS_4His	Derivative of pACYCF1 with 4 His insertion at C-terminus of Caf1M	R1235	This study

2.4 DNA techniques

2.4.1 Plasmid isolation

QIAprep® miniprep kit was used to isolate DNA from cells. There were three steps involved in this procedure, preparation and clearing of bacterial lysate, adsorption of the DNA onto the spin-column membrane and washing and elution of plasmid DNA. Cells from ON cultures (4.5 ml), incubated at 26°C, were pelleted by centrifugation at 13000rpm for 5 minutes and the supernatant completely removed. The bacterial cells were resuspended in 250 µl Buffer P1. P1 buffer contains EDTA which removes Mg²⁺ ion that preserves the cell membrane structure. Then 250 µl of buffer P2 (contain SDS for removing the lipids of the cell membrane) was added and the tube was inverted gently 6 times. At this point, the mixture turned blue. Then, 350 µl of N3 buffer was added to the mixture which was then inverted gently 6 times until the mixture turned colorless. This was to neutralize the lysate and adjust it to a high-salt binding condition that helps renature the DNA and denature the protein, chromosomal DNA, cellular debris and precipitate the SDS. The mixture was centrifuged at 13000 rpm for 10 minutes and the supernatant was applied to a column followed by centrifugation for 60 seconds to bind the DNA to the spin-column membrane. PB buffer (500 µl) was added to the column which was centrifuged for 1 minute at 13000 rpm to ensure efficient

binding of the DNA to the membrane. The supernatant was removed and centrifuged again for 1 minute to remove the rest of the PB buffer from the column. Then, 750 μ l of PE buffer (contains ethanol) was added to the column. The column was placed in a new microcentrifuge tube and 30 μ l sterilized nH_2O was added to the column. After 2 minutes standing at RT, the plasmid DNA was recovered by centrifugation for 1 minute at 13000 rpm. The DNA was then stored at -20°C for later use.

2.4.2 DNA quantitation

The NanoDrop spectrophotometer (Thermo Scientific) was used to determine the concentration of DNA and quality by measuring the ratio of $\text{OD}_{260\text{nm}}/\text{OD}_{280\text{nm}}$. The ratio absorbance at these wavelengths has been used because the maximum absorbance of nucleic acid is at 260 nm and 280 nm for protein. A pure DNA sample will have a ratio of >1.8 $\text{OD}_{260\text{nm}}/\text{OD}_{280\text{nm}}$. A total of 2 μ l DNA was used for quantitation.

2.4.3 Agarose gel electrophoresis

To check the DNA quantity and quality, 0.7% w/v agarose gel was used to separate the DNA fragments. The gel was prepared by melting 0.35 g agarose powder (Melford Labs) in 49.65 ml 1X TAE (40 mM Tris-HCl acetate, 1 mM EDTA pH 8.0). GelRed (Biotium, Cambridge Bioscience) (2.5 μ l or a 1:20000 dilution) was added prior to pouring the gel into a horizontal cast (BioRad) with ten well comb. DNA samples were mixed with 2X final concentration of BlueJuice™ Gel Loading Buffer (Invitrogen) and loaded onto the gel along with 2 μ l of HyperLadder I (Bioline). 1X TAE was used as running buffer, and the gel was run for 90 minutes at 80 Volts. A SYNGENE system (G-Box) was used to image the gels and images were saved in *.tif format.

2.4.4 DNA sequencing and analysis

A total of 5 µl of plasmid DNA per reaction with a concentration of 50-100 ng/µl (from DNA isolation using QIAprep® Miniprep kit) was sent for sequencing with appropriate primer to Source Bioscience, Oxford. If the DNA concentration was lower than 50 ng/µl, the DNA was sent for sequencing with amplification. Primers were by Source Bioscience were sent at a concentration of 3.2 pmol/µl (5µl/reaction). Sequences were analysed using DNAdynamo version 1.374 (Blue Tractor Software Ltd) and FinchTV version 1.4 (Geospiza).

2.5 Mutagenesis and Polymerase Chain Reaction (PCR)

An Eppendorf MasterCycle EP gradient S machine was used to carry out all PCR reactions.

2.5.1 Oligonucleotides

Oligonucleotides for mutagenesis and sequencing primers are listed in Table 2.3 and Table 2.4 respectively. For site-directed mutagenesis; the primer size was between 25 to 45 bases with the desired mutation situated in the middle of the primers and 10 to 15 correct bases on either side. The GC content was ideally close to 40%, the melting temperature at least 78°C (calculated using the Stratagene Quik change formula) and primers terminated with 1-3' C or G bases.

Clontech primer design was used to design primers for inverse PCR of linearized vector and amplification of gene of interest (insert). When designing primers, 15bp of overlapping sequence with adjacent segment was included. This 15bp overlap sequence is crucial for the InFusion enzyme to work by facilitating the complementary

overlapping sequence and create the desire construct. DNAdynamo program was used to predict the T_m and GC content of the primer and to check the plasmid is in frame after cloning. The melting temperature of the primer pair was aimed to be between 58°C to 65°C with a difference of less than 4°C, a GC content of around 40% to 60%, length of 18 to 25 nucleotides (excluding 15bp overlap for infusion) and Gs or Cs at the 3' end. Primers were purchased from Eurofin MWG-Biotech, resuspended in nH₂O to make a stock of 200 pmol/μl, stored at -20°C and used at a working solution of 10 pmol/μl.

Table 2.3: Primers used in the construction of mutants.

Residue in red indicates changes, blue insertion, green substitution, bold indicates restriction site and italic indicates 15bp overlap for infusion.

Primer	Sequence (5'-3')	Comment
Caf1A-F4SFor	GGGACGCGCATATACT TCGGACTCTACTATGC TTG	Mutates Phe4 of Caf1A to Ser. Forward primer.
Caf1A-F4SRev	CAAGCATAGTAGAGTC CGAAGTATATGCGCGT CCC	Mutates Phe4 of Caf1A to Ser. Reverse primer.
Caf1A-Y240AFor	CAGATGAATCGATGGT CCTGCTTACCAATGGA ATTTTGCTCC	Mutates Tyr240 of Caf1A to Ala. Forward primer.
Caf1A-Y240ARev	GGAGCAAATTCATT GGTAGGCAGGAACCAT CGATTCATCTG	Mutates Tyr240 of Caf1A to Ala. Reverse primer.
Caf1A-NdelFor	TTCCCGGGCAATTGTG GGATGACGGCATTCC	Caf1A-ΔN with XmaI restriction site. Forward primer.
Caf1A-NdelRev	ATCCCGGGAGAGTCC GAAGTATATGCGCGTC	Caf1A-ΔN with XmaI restriction site. Reverse primer.
F1for11	CCACTGCAACGGCAAC TCTTG	Amplification of F1D11, F1D12, F1D13, F1::1, F1::2 and F1::3 inserts (Hybridises to <i>caf1</i>). Forward primer.
F1rev11	CTGCTGCAAGTTTACC GCCTTTTG	Amplification of F1D11, F1D12, F1D13, F1::1, F1::2 and F1::3 inserts (Hybridises to <i>caf1</i>). Reverse primer.
InfFor2	AGGCGGTAAACTTGCA GCAGG	Inverse PCR, to linearize pACYCF1 as vector for F1D11, F1D12, F1D13, F1::1, F1::2 and F1::3 (Hybridises to <i>caf1</i>). Forward primer.

InfRev2	ACAAGAGTTGCCGTTG CAGTGG	Inverse PCR, to linearize pACYCF1 as vector for F1D11, F1D12, F1D13, F1::1, F1::2 and F1::3 (Hybridised to <i>caf1</i>). Reverse primer.
F1D14insertFOR-2	GATTTAACTGCAAGCA CCACTGC	Amplification of F1D14 and F1::4 inserts (Hybridised to <i>caf1</i>). Forward primer.
F1D14insertREV-2	GGTTAGATACGGTTAC GGTTACAG	Amplification of F1D14 and F1::4 inserts (Hybridised to <i>caf1</i>). Reverse primer.
F1D14invFOR-2	TGCTGTAACCGTAACC GTATCTAAC	Inverse PCR, to linearize pACYCF1 as vector for F1D14 and F1::4 (Hybridised to <i>caf1</i>). Forward primer.
F1D14invREV-2	TTGCAGTGGTGCTTGC AGTTAAATC	Inverse PCR, to linearize vector pACYCF1 as for F1D14 and F1::4 (Hybridised to <i>caf1</i>). Reverse primer.
Y287AFor-Caf1A	GGTAGAGGTTTAAAGA GATGGC CC ACTGTAA GTAATGAGTTGGTG	Mutates Tyr266 of Caf1A to Ala. Forward primer.
Y287ARev-Caf1A	CACCAACTCATTACTT ACAGTG CC GCCATCTC TTAAAACCTCTACC	Mutates Tyr266 of Caf1A to Ala. Reverse primer.
Y287SFor-Caf1A	GGTAGAGGTTTAAAGA GATGGC AG ACTGTAA GTAATGAGTTGGTG	Mutates Tyr266 of Caf1A to Ser. Forward primer.
Y287SRev-Caf1A	CACCAACTCATTACTT ACAGTG CT GCCATCTC TTAAAACCTCTACC	Mutates Tyr266 of Caf1A to Ser. Reverse primer.
F1ins1_Q190Afor	ACTCTTATTATGCTCAG TTG GC ACCGGGATTAA ACATAGGGG	Mutates Gln169 of Caf1A to Ala to make pACYCF1_A_Q169A and pACYCF1::1_A_Q169A constructs. Forward primer.
F1ins1_Q190Arev	CCCCTATGTTTAATCCC GGT CCA ACTGAGCAT AATAAGAGT	Mutates Gln169 of Caf1A to Ala to make pACYCF1_A_Q169A and pACYCF1::1_A_Q169A constructs. Reverse primer.
F1ins1_Q190Gfor	CTCTTATTATGCTCAGT TG GG ACCGGGATTAAA CATAGGGG	Mutates Gln169 of Caf1A to Gly to make pACYCF1_A_Q169G and pACYCF1::1_A_Q169G constructs. Forward primer.
F1ins1_Q190Grev	CCCCTATGTTTAAT CCC GGTCCCAACTGAGCAT AATAAGAG	Mutates Gln169 of Caf1A to Gly to make pACYCF1_A_Q169G and pACYCF1::1_A_Q169G constructs. Reverse primer.
D12_N80Gfor	CATTTACTTCTCAGGAT GGAG GT AACCAACCAAT TCACTACAAAAGTG	Mutates Asn80 to Gly of Caf1 to make pACYCF1_F1_N80G and pACYCF1::4_F1_N80G constructs. Forward primer.

D12_N80Grev	CACTTTTGTAGTGAATT GGTGGTTA CCT CCATC CTGAGAAGTAAATG	Mutates Asn80 to Gly of Caf1 to make pACYCF1_F1_N80G and pACYCF1::4_F1_N80G constructs. Reverse primer.
D12_N81Gfor	GTACTTAACATTTACTT CTCAGGATGGAAAT GG CCACCAATTCACTAC	Mutates Asn81 to Gly of Caf1 to make pACYCF1_F1_N81G and pACYCF1::4_F1_N81G constructs. Forward primer.
D12_N81Grev	GTAGTGAATTGGTGG C C ATTTCCATCCTGAGA AGTAAATGTTAAGTAC	Mutates Asn81 to Gly of Caf1 to make pACYCF1_F1_N81G and pACYCF1::4_F1_N81G constructs. Reverse primer.
Caf1_P16A_For	CAACGGCAACTCTTGT TGAA G CAGCCCGCATC	Mutates Pro16 to Ala of Caf1 to make pACYCF1_F1_P16A and pACYCF1::3_F1_P16A constructs. Forward primer.
Caf1_P16A_Rev	GATGCGGGCTG C TTCA ACAAGAGTTGCCGTTG	Mutates Pro16 to Ala of Caf1 to make pACYCF1_F1_P16A and pACYCF1::3_F1_P16A constructs. Reverse primer.
Caf1_R18K_For	GCAACTCTTGTTGAAC CAGCC AAG ATCACTCT TACATATAAGGAAG	Mutates Arg18 to Lys of Caf1 to make pACYCF1_F1_R18K and pACYCF1::3_F1_R18K constructs. Forward primer.
Caf1_R18K_Rev	TTCCTTATATGTAAGA GTGAT CTT GGCTGGTT CAACAAGAGTTGCC	Mutates Arg18 to Lys of Caf1 to make pACYCF1_F1_R18K and pACYCF1::3_F1_R18K constructs. Reverse primer.
F1ins2_ELDKWANH_for	CTGAACTGGATAAATG GGCC AACCACCA ATTC ACTACAAAAGTGATTG	Inserts His and Asn to create pACYCF1::2_ELDKWANH construct. Forward primer.
F1ins2_ELDKWANH_rev	CAATCACTTTTGTAGT GAAT TTGGT GTTGGCC CATTTATCCAGTTCAG	Inserts His and Asn to create pACYCF1::2_ELDKWANH construct. Reverse primer.
Caf1A_S203A_for	CTTGGCGCTTTCGTAGT G CAACCTCATGGTGGA AA	Mutates Ser182 to Ala of Caf1A to make pACYCF1_A_S182A and pACYCF1::1_A_S182A constructs. Forward primer.
Caf1A_S203A_rev	TTTCCACCATGAGGTT G CACTACGAAAGCGCC AAG	Mutates Ser182 to Ala of Caf1A to make pACYCF1_A_S182A and pACYCF1::1_A_S182A constructs. Reverse primer.
Caf1A_Q188A_for	TCTCTGGACTCTTATTA TGCT GC GTTGCAACCG GGATTAAAC	Mutates Gln167 of Caf1A to Ala. Forward primer.
Caf1A_Q188A_rev	GTTTAATCCCGGTTGC AAC GC AGCATAATAAG AGTCCAGAGA	Mutates Gln167 of Caf1A to Ala. Reverse primer.
Caf1A_Q188S_for	GTCTCTGGACTCTTATT ATGCT TC GTTGCAACC GGGATTAAACA	Mutates Gln167 of Caf1A to Ser. Forward primer.

Caf1A_Q188S_rev	TGTTTAATCCCGGTTGC AAC GA AGCATAATAAG AGTCCAGAGAC	Mutates Gln167 of Caf1A to Ser. Reverse primer.
F1D12_GGGGNNH_F OR	CTGGTGGGGGTGGC AA TAACCACCA ATTCACT ACAAAAGTG	Inserts His and two Asn to create pACYCF1D12_ELDKWANNH construct. Forward primer.
F1D12_GGGGNNH_R EV	CACTTTTGTAGTGAATT G GTGGTT A TTGCCACC CCCACCAG	Inserts His and two Asn to create pACYCF1D12_ELDKWANNH construct. Reverse primer.
F1ins2_ELDKWANN H_for	CTGAACTGGATAAATG GGCC AATAACCACCA TTCACACAAAAGTG	Inserts His and two Asn to create pACYCF1::2_ELDKWANNH construct. Forward primer.
F1ins2_ELDKWANN H_rev	CACTTTTGTAGTGAATT G GTGGTT A TTGGCCCA TTTATCCAGTTCAG	Inserts His and two Asn to create pACYCF1::2_ELDKWANNH construct. Reverse primer.
Plug_E298K_For	GAGCTGAAAGTCATCA TTCAT AAA AGTGATGG AACAAAGCAAG	Mutates Glu298 of Caf1A middle domain to Lys. Forward primer.
Plug_E298K_Rev	CTTGCTTTGTTCCATCA CTTT T ATGAATGATGA CTTTCAGCTC	Mutates Glu298 of Caf1A middle domain to Lys. Reverse primer.
Plug_E291K_For	GGGTGGGGGGAGTGGT A AGCTGAAAGTCAT	Mutates Glu291 of Caf1A middle domain to Lys. Forward primer.
Plug_E291K_Rev	TGACTTTCAGCT T ACC ACTCCCCCACCC	Mutates Glu291 of Caf1A middle domain to Lys. Reverse primer.
Plug_E291A_For	GTGGGGGGGAGTGGT GC GCTGAAAGTCATCAT	Mutates Glu291 of Caf1A middle domain to Ala. Forward primer.
Plug_E291A_Rev	ATGATGACTTTCAG CG CACCACCTCCCCCAC	Mutates Glu291 of Caf1A middle domain to Ala. Reverse primer.
Plug_R263A_For	GCCAGGGTAGAGGTTT TAG CA AGATGGCTACAC TGTAAG	Mutates Arg263 of Caf1A middle domain to Ala. Forward primer.
Plug_R263A_Rev	ACTTACAGTGTAGCCA TCT CA TAAACCTCTA CCCTGGC	Mutates Arg263 of Caf1A middle domain to Ala. Reverse primer.
Hairpin_Y215A_F	CGTTTGACATTGGGGG AAACC GCT TCTGATAG CAGTATCTTTG	Mutates Tyr215 of Caf1A hairpin loop to Ala. Forward primer.
Hairpin_Y215A_R	CAAAGATACTGCTATC AGAA CGCG TTTCCCCC AATGTCAAACG	Mutates Tyr215 of Caf1A hairpin loop to Ala. Reverse primer.
Hairpin_D217A_F	GGGGGAAACCTATTCT GCT AGCAGTATCTTTG ACAG	Mutates Asp217 of Caf1A hairpin loop to Ala. Forward primer.
Hairpin_D217A_R	CTGTCAAAGATACTGC TAG CA GAAATAGGTTTC CCCC	Mutates Asp217 of Caf1A hairpin loop to Ala. Reverse primer.
Plug_S289A_F	CTCTGGGTGGGGGG GC TGGTGAGCTGAAAG	Mutates Ser289 of Caf1A middle domain to Ala. Forward primer.
Plug_S289A_R	CTTTCAGCTCACC AGC CCCCCACCCAG	Mutates Ser289 of Caf1A middle domain to Ala. Reverse primer.

Plug_R258A_F	CGCACGTACACAAGCC GCGGTAGAGGTTTTAA GAG	Mutates Arg258 of Caf1A middle domain to Ala. Forward primer.
Plug_R258A_R	CTCTTAAAACCTCTAC C GCGGCTTGTGTACGT GCG	Mutates Arg258 of Caf1A middle domain to Ala. Reverse primer.
Barrel_K462A_F	GATTATGCTAAACCCA AAAACG CAGTGCAATT CAATTTAAGTCAAAGC	Mutates Lys462 of Caf1A β -barrel to Ala. Forward primer.
Barrel_K462A_R	GCTTTGACTTAAATTG AATTGCACTG C G TTTT GGGTTTAGCATAATC	Mutates Lys462 of Caf1A β -barrel to Ala. Reverse primer.
Barrel_N144A_F	GGCATTCCTGCTCTGTT CATG C T TATAATACG AACATGCAGAC	Mutates Asn144 of Caf1A β -barrel to Ala. Forward primer.
Barrel_N144A_R	GTCTGCATGTTCTGATT ATAA G C CATGAACAGA GCAGGAATGCC	Mutates Asn144 of Caf1A β -barrel to Ala. Reverse primer.
Barrel_N148A_F	GCTCTGTTTCATGAATT ATAATACG C CATGCA GACAAGAAAATTCAG	Mutates Asn148 of Caf1A β -barrel to Ala. Forward primer.
Barrel_N148A_R	CTGAATTTTCTTGTCTG CATG C C G TATTATAA TTCATGAACAGAGC	Mutates Asn148 of Caf1A β -barrel to Ala. Reverse primer.
Barrel_Q150A_F	GAATTATAATACGAAC ATG C G GACAAGAAAAT TCAGAGAAGGAGGC	Mutates Gln150 of Caf1A β -barrel to Ala. Forward primer.
Barrel_Q150A_R	GCCTCCTTCTCTGAATT TTCTTG T C GCCATGTTC GTATTATAATTC	Mutates Gln150 of Caf1A β -barrel to Ala. Reverse primer.
Barrel_Q464A_F	GCTAAACCCAAAAACA AAGTG C CATTCAATTT AAGTCAAAGCATACCT G	Mutates Gln464 of Caf1A β -barrel to Ala. Forward primer.
Barrel_Q464A_R	CAGGTATGCTTTGACT TAAATTGAAT C C CACT TTGTTTTGGGTTTAGC	Mutates Gln464 of Caf1A β -barrel to Ala. Reverse primer.
Barrel_N601A_F	CCTCCAATGCACTGAA TTTG C T TATCGAGGA ACTTATGGG	Mutates Asn601 of Caf1A β -barrel to Ala. Forward primer.
Barrel_N601A_R	CCCATAAGTTCCT C G A TAAGCCAAATTCAGTG CATTGGAGG	Mutates Asn601 of Caf1A β -barrel to Ala. Reverse primer.
SE11For1	AGGAGATATACCATGA ACGGCAGCTATGCGC CG	Amplification of <i>E. coli</i> SE11 MA_S123 inserts (Hybridises to pACYCD1). Forward primer. To create pACYC_MA_S123 construct.

SE11Rev1	<i>TGGCTGCTGCCCATG</i> <i>TGTAAGAACATGCCT</i> <i>GCAC</i>	Amplification of <i>E. coli</i> SE11 MA_S123 inserts (Hybridises to pACYCD1). Reverse primer. To create pACYC_MA_S123 construct.
pACYCD1_MA_S1_F	<i>GAACTAGTTTATATAC</i> <i>CAAATGAAAGC</i>	SE11_ΔS23 with SpeI restriction site. Forward primer.
pACYCD1_MA_S1_R	<i>GAACTAGTAGCAGCC</i> <i>ATCACCATC</i>	SE11_ΔS23 with SpeI restriction site. Reverse primer.
SE11_Mhis_C2_F	<i>AGGAAACAGACCATGAT</i> <i>GAAGTATAGAAAATA</i> <i>CAAATCTTAAATTGTG</i>	Amplification of <i>E. coli</i> SE11 M6his_S123 inserts (Hybridises to pACYCD1). Forward primer. To create pTRC99_Mh_S123 construct.
SE11_Mhis_C2_R	<i>GAACACCATTAGTGATG</i> <i>ATGATGATGATGAAAA</i> <i>ATTACATTCTTCG</i>	Amplification of <i>E. coli</i> SE11 M6his_S123 inserts (Hybridises to pACYCD1). Reverse primer. To create pTRC99_Mh_S123 construct.
pTRC99_Mh_S1_for	<i>GAACTAGTTTATATAC</i> <i>CAAATGAAAGC</i>	SE11_ΔS23 with SpeI restriction site. Forward primer.
pTRC99_Mh_S1_rev	<i>GAACTAGTTTCTAACG</i> <i>AAGTTTATTGG</i>	SE11_ΔS23 with SpeI restriction site. Reverse primer.
pTRC99_Mh_S23_for	<i>GAACTAGTACTTCGTT</i> <i>AGAAATAAGG</i>	SE11_ΔS1 with SpeI restriction site. Forward primer.
pTRC99_Mh_S23_rev	<i>GAACTAGTCAGACGAT</i> <i>AACGAGAC</i>	SE11_ΔS1 with SpeI restriction site. Reverse primer.
pTRC99-Mh-S3_for	<i>GAACTAGTCAGGGGA</i> <i>ATGATTGTTC</i>	SE11_ΔS12 with SpeI restriction site. Forward primer.
pTRC99-Mh-S3_rev	<i>GAACTAGTCAGACGAT</i> <i>AACGAGAC</i>	SE11_ΔS12 with SpeI restriction site. Reverse primer.
SE11_M8his_for	<i>CGAAGAATGTAATTTT</i> <i>TCACCACCATCATCAT</i> <i>CATCATCAC</i>	Inserts two His of pTRC99-Mh-S123 to create pTRC99-M8h-S123 construct. Forward primer.
SE11_M8his_rev	<i>GTGATGATGATGATGA</i> <i>TGGTGGTGAAAAATTA</i> <i>CATTCTTCG</i>	Inserts two His of pTRC99-Mh-S123 to create pTRC99-M8h-S123 construct. Reverse primer.
F1ins1_6His_For	<i>GTGATTGGCAAGCATC</i> <i>ATCACCATCACCACGAT</i> <i>ATCTCTCCTAAG</i>	Substitutes ELDKWA of pACYCF1::1 to six His to create pACYCF1::1_6His construct. Forward primer.
F1ins1_6His_Rev	<i>CTTAGGAGAGATATCG</i> <i>TGGTGATGGTGATGAT</i> <i>GCTTGCCAATCAC</i>	Substitutes ELDKWA of pACYCF1::1 to six His to create pACYCF1::1_6His construct. Reverse primer.
Fi1ns1_DS4his_F	<i>GATTGGCAAGGATTCT</i> <i>CATCATCACCACGATA</i> <i>TCTC</i>	Substitutes two His of pACYCF1::1_6His to Asp and Ser to create pACYCF1::1_DS_4His construct. Forward primer.

Fi1ns1_DS4his_R	GAGATATCGTGGTGAT GATGAGAAATCCTTGCC AATC	Substitutes two His of pACYCF1::1_6His to Asp and Ser to create pACYCF1::1_DS_4His construct. Reverse primer.
------------------------	---	---

Table 2.4: Sequencing primers

Primer	Sequence (5'-3')	Direction	Comment
DuetDOWN1	GATTATGCGGCCGT GTACAA	Reverse	Primes within vector, downstream Caf1, available commercially
DuetUP1	GTCCGGGATCTCGA CGCTCTCCCTTATG	Forward	Primes within vector, downstream Caf1R, available commercially
F1#1	ATGTTGGGTCTGAAC ATAAATCG	Forward	Primes within vector, near Caf1R
F1#4	CACAAACATCGTAT GTTGGCGCTC	Forward	Primes from middle of Caf1A
F1#6	CCTTCTGCTGAATCC TGAATAC	Forward	Primes from beginning of Caf1R towards intergenic promoter region
F1-7for	TCAGAGGATCCTTT CGTGGTCAC	Forward	Primes from middle of Caf1M
Caf1Afor-3b	GAGGAACTTATGGG GAGATCAGTG	Forward	Primes from end of Caf1A towards Caf1
Caf1Afor-4	AAATCTGGTACAGA GCAATGTG	Forward	Primes from start of Caf1A
Caf1Arev-1	CTACGAAAGCGCCA AGCCCCTATG	Reverse	Primes from within first quarter of Caf1A towards Caf1M
F1rev2	GATTGGTGTGCTG GGTATG	Reverse	Primes from start of intergenic promoter region towards Caf1R
Caf1Arev1	CATAGGGGCTTGGC GCTTTCGTAG	Reverse	To sequence pACYCF1-Caf1A -F4S and -NΔ
Caf1Arev2	CCCAGTTACTTTACC CAATGATGCAGAG	Reverse	To sequence pACYCF1-Caf1A- Y240A
pTRC99A.seq.for	TGCAGGTCGTAAAT CACTGC	Forward	Primes within vector, upstream chaperone of pTRC99_Mh_S123. Available commercially
pTRC99his reverse	CTTCTGCGTTCTGAT TTAATCTG	Reverse	Primes within vector, downstream subunit 3 of pTRC99_Mh_S123. Available commercially

SE11_A_1_Rev	ACACGGTGATAATG TATCTTCG	Reverse	Primes from beginning of usher towards subunit 1
SE11_A_1_For	AATGGCGAAGATAC ATTATCACCG	Forward	Primes from beginning of usher towards chaperone
SE11_A_2_For	AAATTTGATACACG TATCGGTGGG	Forward	Primes from end of usher towards subunit 1
SE11_A_2_Rev	TAACAGTCGCCCAC CGATAC	Reverse	Primes from end of usher towards chaperone
SE11_S2_3_Rev	TAGCAGTTATCGCT GTTCCG	Reverse	Primes from beginning of subunit 2 towards subunit 3
SE11_S2_3_for	TTATTCGGAACAGC GATAACTGC	Forward	Primes from beginning of subunit 2 towards subunit 1

**** See Appendix 13 for approximate location of priming site**

2.5.2 Site-directed mutagenesis

QuikChange II XL Site-Directed Mutagenesis was used to construct all specific mutant of Caf1 and Caf1A in pACYCF1 and Histidine insertions of pTRC99-Mh-S123, pTRC99-M8h-S123, pACYCF1::1_DS_6His and pACYCF1::1_DS_4His. The reaction mixture was prepared, according to the kit instructions and as follows: 2.5 µl 10X reaction buffer, 50 ng of template plasmid, 0.625 µl of 10 pmol forward and reverse plasmid, 0.5 µl of dNTP mix, 1.5 µl of Quick solution agent, 0.5 µl of Pfu Ultra HF DNA polymerase (2.5U/µl) and nH₂O to make up the final volume to 25 µl. The reaction mixture was prepared on ice. The PCR cycling parameter were as follows: initial denaturation at 95°C for 2 minutes (1 cycle), 18 cycles of denaturation (95°C for 1 minute), annealing (60°C for 1 minute) and extension (95°C for 2 min/Kb) followed by final extension at 68°C for 7 minutes. The PCR products were treated with DpnI enzyme from the kit as follows: 0.5 µl of DpnI was added to 15 µl of PCR reaction, incubated for 2 hours at 37°C and transformed into *E. coli* competent cells.

Transformants were picked and grown overnight at 26°C for plasmid isolation and sequencing.

2.5.3 Deletion mutagenesis

Inverse PCR was used for construction of the N-terminal deletion mutant, pACYCF1-A-ΔN. Oligonucleotides (Caf1A-N-delFor and Caf1A-delRev) were designed with a XmaI restriction site, for religation of the PCR product. The PCR mixture was prepared according to CloneAMPTM HiFi Premix instructions. The mixture was prepared in triplicate for three different annealing temperatures to get optimum PCR products using the temperature gradient programme of the PCR machine. Then, the PCR product was treated with DpnI, purified and digested with XmaI, (New England Biolab) to create sticky ends. Then, the restriction enzyme was heat inactivated at 20 minutes at 65°C and the PCR product purified. To ligate the product, DNA (50ng) was made up to 10μl with nH₂O, mixed with Quick ligase (Biolabs) (1 μl) and 10 μl of 2X Quick Ligase buffer (Biolabs) and incubated at room temperature (25°C) for 5 minutes. The plasmid was transformed into *E. coli* DH5α and incubated at 26°C for 48 hours. Transformants were picked and grown overnight at 26°C for plasmid isolation and sequencing.

Table 2.5: PCR reaction mixture of inverse PCR

Component	Volume
Clone Amp HiFi PCR Premix 2X (Clonetech)	12.50 μl
Primer Caf1A-NdelFor (7.5 pmol)	0.75 μl
Primer Caf1A-NdelRev (7.5 pmol)	0.75 μl
pACYCF1 template (50 ng)	5 μl
nH ₂ O	6 μl
Total	25 μl

Table 2.6: PCR reaction condition for inverse PCR.

Different annealing temperatures were used to get optimum PCR product. Annealing temperatures were chosen based on annealing temperature of the oligonucleotides. Below is an example of PCR condition used to synthesize Caf1A-N-terminal deletion.

Step	Temperature	Time	Cycle
Denature	98°C	10 seconds	30 cycles
Annealing	55.1°C, 57.1°C and 59.7°C	10 seconds	
Extension	72°C	88 seconds (10 sec/kb)	

2.5.4 InFusion Mutagenesis

Plasmid constructs pACYC-F1D11, pACYC-F1D12, pACYC-F1D13, pACYC-F1D14, pACYC-F1::1, pACYC-F1::2, pACYC-F1::3, pACYC-F1::4, pACYC_MA_S123 and pTRC99-Mh-S123 were designed using In-fusion HD cloning kit (Clontech). The CloneAMPTM HiFi Premix was used to linearise vectors and amplify inserts.

2.5.4.1 Linearizing vector

Inverse PCR or restriction enzyme digestion was used to linearize vector for infusion. Inverse PCR was used to linearise vector and prepared according to the CloneAMPTM HiFi Premix instructions as follows: 12.5 µl of 2X CloneAmp HiFi PCR Premix, 7.5 µl of 10 pmol forward and reverse primer, 50 and 100 ng of template plasmid, made up to 25 µl with nH₂O. To get optimal PCR products, two different template concentrations were prepared. Cycling parameters with three different

annealing temperatures were used based on primer melting temperature, which were between 2°C to 5°C less than melting temperature. The PCR cycling parameters are as follows: denature for 10 seconds at 98°C, annealing for 10 seconds at three different annealing temperatures, extension at 72°C for 30 secs/Kb.

For linearising vector by restriction digestion, 1µl of restriction enzyme (New England Biolabs) and 10X NEBuffer was added to 30µl of DNA. Then, nH₂O was added up to a final concentration of 50µl. The mixture was incubated for 1 hour at room temperature. The reaction was stopped by heat inactivation according to manufacturer instruction (i.e. SpeI was heat inactivated at 80°C for 20 minutes).

After amplification, the linearised vector was verified using agarose gel, DpnI treated and purified. The purified linerised vector and insert were then analysed again by agarose gel to check the quality and quantity of the purified PCR products.

2.5.4.2 Amplification of inserts

CloneAMPTM HiFi Premix was used to amplify the targeted insert. The PCR reaction mix was prepared as follows: 12.5 µl of 2X CloneAmp HiFi PCR Premix, 7.5 µl of 10 pmol forward and reverse primer, 50 ng of template plasmid, made up to 25 µl with nH₂O. To get optimal PCR products, cycling parameters with three different annealing temperatures were used based on primer melting temperature which are between 2°C to 5°C less than melting temperature. The PCR cycling parameter are as follows: denature for 10 seconds at 98°C, annealing for 10 seconds at three different annealing temperatures, extension at 72°C for 30 secs/Kb. Then, the PCR product was DpnI treated, purified and quality and quantity of the product was checked by agarose gel.

2.5.4.3 InFusion cloning reaction

InFusion cloning reactions were set up at room temperature as follows: 2 μ l 5X InFusion enzyme premix, 1 volume of vector (100ng), 3 volume of purified insert, made up with nH_2O to final volume of 10 μ l. In-fusion molar ratio calculator was used to help calculate the optimal amount of vector and insert for the In-Fusion. (http://www.clontech.com/US/Products/Cloning_and_Compotent_Cells/Cloning_Resources/Online_In-Fusion_Tools). The mixture was incubated for 15 minutes at 50°C, and then placed on ice. Finally, 3 μ l of the reaction mixture was used for transformation of Stellar competent cells (Clontech) which were incubated for 48 hours at 26°C.

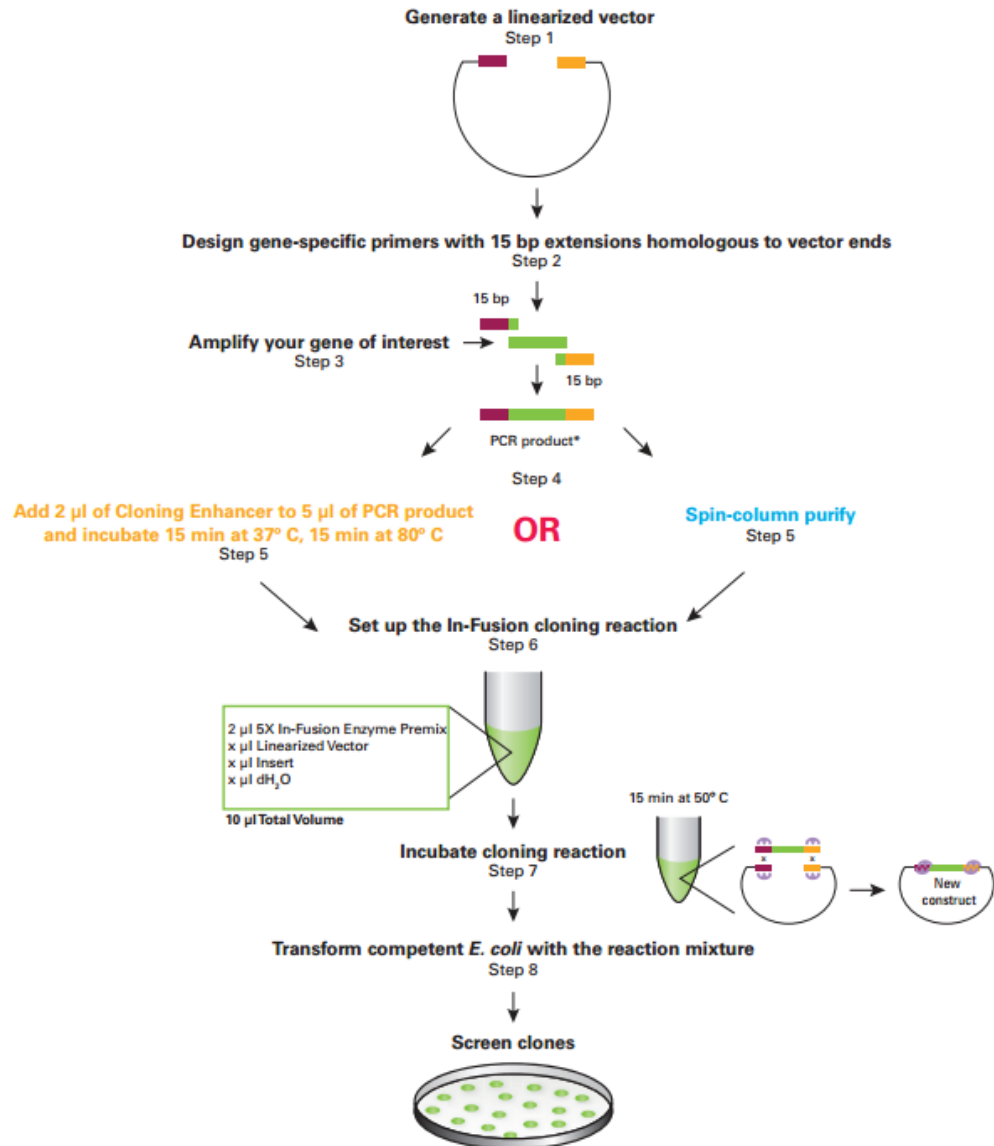


Figure 2.1: In-Fusion cloning procedure using In-Fusion HD Cloning kit (Image taken from Clontech In-Fusion user manual)

2.5.5 DpnI digestion

Following PCR, DpnI digestion was used to digest methylated and hemimethylated parental template in the reaction mixture. DpnI CutSmart (0.5µl)(New England Biolabs) was added to 21µl of the PCR product, 1X CutSmart buffer (New England Biolabs) and nH₂O was added up to final concentration of 25µl. The mixture was incubated for 1 hour at 37°C.

2.5.6 Purification of PCR products

QIAquick® PCR Purification Kit was used to purify the PCR products of 100bp-10Kbp. Sample (usually all PCR product was purified, 25µl) was mixed with 5 volumes of PB buffer. This buffer helps to increase the negative charge of the DNA so that it could bind to the column. The mixture was then applied to a QIAquick column and centrifuged for 60 seconds. PE buffer (750 µl), which contains ethanol, was added to the column and centrifuged for another 60 seconds to wash the DNA. Then the column was placed in a clean 1.5 ml microcentrifuge tube. Sterilised nH₂O (50 µl) was added to the column and centrifuged to elute the DNA. A Nanodrop spectrophotometer was used to quantify the purified DNA which was then stored at -20°C for later use.

For infusion, inserts and vectors were purified using NucleoSpin® Gel and PCR Clean-up kit (MACHERY-NAGEL). The PCR product was mixed with 2 volumes of buffer NT1 (binding buffer). The mixture was applied onto a NucleoSpin column and centrifuged for 30 seconds at 10000 rpm. The flow-through was discarded. Then, the silica membrane was washed by adding 700µl Buffer NT3 (ethanolic buffer) to the column followed by centrifugation for 30 seconds at 10000 rpm. The flow-through was

discarded and the silica membrane was dried by centrifugation for 1 minute at 11000 x g. Pure DNA was eluted with 30µl Buffer NE, incubated at RT for 1 minute and centrifuged for 1 minute at 10000 rpm. The DNA concentration was read using Nanodrop spectrophotometer then stored at -20°C for later use.

2.5.7 DNA extraction from agarose gels

NucleoSpin® Gel and PCR Clean-up kit (MACHERY-NAGEL) was used to extract DNA from agarose gel. After running 0.7% agarose gel for 90 minutes at 80 Volts, a clean scalpel was used to excise the DNA fragment. All excess agarose was removed and weight of the gel slice was determined. Based on the weight, for each 100 mg of agarose gel, 200µl Buffer NTI (binding buffer) was added and incubated for 5 to 10 minutes at 50°C until the gel slices completely dissolved. The dissolved gel was transferred onto a NucleoSpin column, centrifuged for 30 seconds at 10000 rpm. The flow-through was discarded. Then, the silica membrane was washed by adding 700µl Buffer NT3 (ethanolic buffer) to the column followed by centrifugation for 30 seconds at 10000 rpm. The flow-through was discarded and the silica membrane was dried by centrifugation for 1 minute at 10000 rpm. Pure DNA was eluted with 30µl Buffer NE, incubated at RT for 1 minute and centrifuged for 1 minute at 10000 rpm. The DNA concentration was read using a Nanodrop spectrophotometer then stored in -20°C for later use.

2.6 Making competent cells

An overnight culture (0.5 ml) was subcultured into 50ml LB in 250 ml flask and incubated at 37°C with 225 rpm shaking until the OD₆₀₀ reached 0.4-0.5, at which point

cells were harvested in Falcon tubes by centrifugation at 4000 rpm and 4°C for 15 minutes. The supernatant was discarded and the pellet was resuspended in 15 ml ice-cold RF1 buffer. The mixture was incubated for 15 minutes and then centrifuged at 4000 rpm for 10 minutes at 4°C. The pellet was resuspended in 5ml ice-cold RF2 buffer and incubated on ice for 15 minutes. Cells were dispensed in 50 µl aliquots in sterile microcentrifuge tubes, using cold tips. Finally, the aliquots were snap frozen in liquid N₂ and stored at -80°C.

Table 2.7: Preparation of RF1 and RF2 buffer.

RF1 buffer composition (160mM KCl, 50mM MnCl ₂ , 30 mM K acetate, 10 mM CaCl ₂ , 15% glycerol, pH 5.8)		RF2 buffer composition (10 mM KCl, 75 mM CaCl ₂ , 10 mM MOPS, 15% glycerol, pH 6.8)	
KCl	2.4 g	KCl	0.3 g
MnCl₂.4H₂O	1.98 g	0.5 M MOPS (pH 6.8 adjust with KOH)	4 ml
Potassium Acetate	0.6 g		
Glycerol	30 ml	Glycerol	30 ml
CaCl₂.2H₂O	0.3g	CaCl₂.2H₂O	2.2 g
Adjust pH 5.8 with 0.2M acetic acid, make up to 200 ml and filter sterilise. Store at -20°C.		Adjust pH 6.8 with KOH, make up volume to 200ml and filter sterilise. Store at -20°C.	

2.7 Transformation

Competent cells from -80°C stocks were thawed on ice. DNA from plasmid isolation (1 µl), InFusion/ligation mixture (3 µl) or site-directed mutagenesis (1 or 2 µl) was added and mixed gently with 25 µl or 30 µl competent cells, respectively. The mixture was incubated for 30 minutes on ice, heat-shocked at 42°C on a heating block for 1 minute, and then incubated for 2 minutes on ice. Pre-warmed LB (250 µl) with 0.6% glucose was added to the cells, which were then incubated at 37°C for 1 hour at 225 rpm shaking. Transformants were spread on LB agar plates with the appropriate

antibiotic and incubated for 48 hours at 26°C, inverted. Transformants were stored on plates at -4°C until required (generally within 1 or 2 days, within 2 weeks at latest). For storage, single transformants were picked and resuspended in 1 ml of LB with 15% glycerol. The transformant stock was dispensed into 100 µl aliquots and stored at -80°C for subsequent experiments.

2.8 Cell fractionation

2.8.1 Analysis of F1 in whole *E. coli* cells

E. coli DH5α or *E. coli* BW25113 carrying pACYCF1 or derivative was cultured overnight and induced at 37°C as in section 2.2. After 4 h induction, OD₆₀₀ was measured. Then, 1 ml of cells was centrifuge at 13000 rpm for 20 minutes. Encapsulated cells require a longer time to pellet than unencapsulated cells. The supernatant was discarded by pipetting and avoiding losing any of the fluffy pellet (F1). The pelleted cells were resuspended in PBS to give a final concentration of 5OD cell equivalent/ml and stored at -20°C.

2.8.2 Isolation of surface F1

E. coli DH5α or *E. coli* BW25113 carrying pACYCF1, derivative thereof, or pFMA1 was cultured overnight and induced for F1 production, as in section 2.2. Following induction, 3OD equivalent of cells were recovered by centrifugation at 13000 rpm for 20 minutes. The pelleted cells were resuspended in 100 µl PBS and heat-shocked for 1 hour at 57°C, in a heating block or water bath unless otherwise indicated. The cells were mixed gently every 15 minutes. Then, the cells were removed at 13000

rpm for 10 minutes at 4°C. The supernatant that contained released F1 was stored at -20°C.

Preparation of the F1 standard used the same procedure as surface extraction except it was done in batch. First, the DH5α containing pACYCF1 plasmid was cultured in 10 ml LB, 0.6% glucose, 34 µg/ml CAM at 26°C ON with 225 rpm shaking. The next day, 10 ml of the washed cells were subcultured in 100 ml of LB with 10 µg/ml CAM and incubated for 4 hours at 37°C with 225 rpm shaking, then the OD600 was read. All of the cells from 100 ml LB were recovered by centrifugation at 15000 rpm for 30 minutes at 4°C using the Sorvall high-speed centrifuge. The recovered cells were resuspended in PBS (1ml of PBS per 10 ODs bacterial cell) and heat-shocked for 1 hour at 57°C in a water bath. The cells were mixed gently every 15 minutes. Then the sample was centrifuged at 15000 rpm for 10 minutes at 4°C using the Sorvall centrifuge. The supernatant was kept at -20°C and quantified to establish a value for the standard.

2.8.3 Isolation of periplasmic F1

E. coli DH5α or *E. coli* BW25113 carrying pACYCF1 or pFMA1 were cultured overnight and thermo-induced at 37°C for 4h to induce for F1 production, or with IPTG as in section 2.2. Following induction, 3OD cells were recovered by centrifugation at 13000 rpm for 20 minutes. The pelleted cells were resuspended in 100 µl of 20% (w/v) sucrose, in 30mM Tris pH 8.0, 10mM EDTA and incubated for 10 minutes at RT. Plasmolysed cells were recovered by centrifugation at 13000 rpm for 10 minutes at RT. All supernatant was removed and cells resuspended in 100 µl of cold 10mM MgCl₂ and incubated for 10 min on ice. The periplasmic fractions were recovered by centrifugation

at 13000 rpm for 10 minutes at 4°C. The supernatant (periplasmic fractions) was stored at -20°C for further analysis.

2.9 Protein techniques

2.9.1 Bradford assay

A Quick Start Bradford Protein Assay kit (Bio-Rad) was used to determine the concentration of the Caf1 protein in the Caf1 standard. The Microassay protocol was used with bovine gamma-globulin as standard (1.25-20µg) prepared as shown in Table 2.8. Duplicates of 150 µl of each concentration of the standard were pipetted into microplate wells and mixed with 150 µl of 1X dye reagent. At the same time, samples of the F1 standard (unknown concentration) were also prepared as shown in Table 2.9. The mixture was then incubated for 15 minutes at RT before being measured using the microplate reader at an absorbance of 595 nm. A standard curve of the gamma-globulin was plotted to determine the concentration of F1 standard.

Table 2.8: Preparation of microplate gamma-globulin standard dilutions, sufficient for duplicates. 150 μ l of each dilution is used in microassay.

Tube	Standard volume (μ l)	Source standard	of Diluent volume / PBS (μ l)	Final protein concentration (μ g/ml)
1	10	2 mg/ml stock	790	25
2	10	2 mg/ml stock	990	20
3	6	2 mg/ml stock	794	15
4	500	Tube 2	500	10
5	500	Tube 4	500	5
6	500	Tube 5	500	2.5
7	500	Tube 6	500	1.25
8 (blank)	-	-	500	0

Table 2.9: Preparation of microplate samples of F1.

The F1 samples were prepared directly in microtitre plate as shown (150 μ l final volume). Dilutions prepared and run in duplicate.

F1 standard applied (μ l)	Diluent volume / PBS (μ l)
10	140
8	142
6	144
4	146
2	148

2.9.2 Quantitation of Caf1 from Coomassie blue stained SDS-PAGE

Samples of heat extracted F1 were run in duplicate or triplicate (as indicated) Mini-PROTEAN® TGX™ precast gels (Bio-Rad). All gels for one experiment were stained (1 hour) and destained (45minutes with two changes of destain) at the same time in the same plastic container. A SynGene (G-box) was used to image the gels and GeneTools program was used for quantitation by using manual quantitation option. A circular drawn and adjusted to match the size of each band. Then, the background was

subtracted by using automatic background correction. The raw data of each band from the gel was averaged for comparison of mutants; wild-type was assigned as 100%.

2.9.3 Sodium-dodecyl-sulphate polyacrylamide gel electrophoresis

A mini-PROTEAN II gel apparatus (Bio-Rad), which is a vertical electrophoresis system, was used to resolve the protein sample using sodium-dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE) (Laemmli, 1970). Samples were mixed with 4X sample buffer (62.5 mM Tris HCL, 2% SDS, pH 6.8, 10% glycerol, 0.0024% bromophenol blue) and boiled for 15 minutes at $>95^{\circ}\text{C}$ to denature the Caf1 polymer or used unheated to analyse the Caf1 polymer. The concentrations of the gels were 14% acrylamide for separating and 5% acrylamide for stacking gels. Gels used were 0.75 mm thick and were run at 200 V for 55 minutes with freshly diluted SDS-PAGE running buffer (0.25 M Tris HCL, 1.92 M Glycine pH 8.3, 1% SDS). Finally, the gels were stained with Coomassie Blue (2.5 g Coomassie Blue R250, 450 ml methanol, 450 ml H₂O, 100 ml acetic acid) for 1 hour to overnight followed by destaining for 1 hour with multiple changes of de-stain solution (30% methanol, 10% glacial acetic acid, 60% dH₂O). For quantitation, SDS-PAGE gels were stained (1 hour) and destained (45 minutes with two changes of destain) at the same time. A SynGene (G-box) system was used to image the gels.

2.9.4 Immunoblotting

Hybond ECL nitrocellulose membrane (Life Sciences); 0.45 μm and a Mini-PROTEAN II Mini Trans-blot module (Bio-Rad) with transfer buffer (0.025 M Tris-HCL, 0.2 M glycine) were used to identify protein separated by SDS-PAGE. Following

blotting, the membrane was blocked overnight at 4°C in 50 ml blocking buffer (Tween 20 Tris buffer saline (TTBS, 0.05M Tris (pH 7.5), 0.15M NaCl, 0.1% (v/v) Tween 20) containing 5% milk powder). The next day, the membrane was incubated for a further 30 minutes at 25°C prior to washing. The membrane was rinsed twice in TTBS followed by three 5 minutes washes in TTBS. Primary antibody was added and the membrane incubated for 1 hour rocking at 25°C, using rabbit polyclonal Caf1SC (14685), to detect Caf1 at a dilution of 1 in 10000 in TTBS or rabbit polyclonal Caf1M at a dilution of 1 in 20,000 to detect Caf1M. After 1 hour, the membrane was rinsed and washed as stated before and secondary antibody (anti-rabbit IgG-peroxidase, Sigma-Aldrich) was added at 1 in 40000 dilutions and the membrane incubated for another 1 hour, rocking at 25°C. Following a quick rinse and three short washes (5 minutes), bound antibody was visualised using a ChemiFast chemiluminescence substrate from Syngene. ChemiFast luminol (500µl) and ChemiFast stable peroxide solution (500µl) were mixed and poured evenly over the membrane. After 5 minutes incubation, the G-Box was used for imaging.

2.9.5 TCA precipitation

For TCA precipitation, surface extraction procedure was followed. Cells were grown overnight at 26°C with 0.6% glucose and appropriate antibiotic, washed, subculture 1/10 with appropriate antibiotic and induced for 4 hours at 37°C. After induction, 1 ml culture was centrifuged for 20 minutes. The supernatant was kept for TCA precipitation. Then, a total of 10% TCA (40% w/v stock) was added to the protein sample and incubate on ice for 1 hour followed by centrifugation for 15 minutes at 4°C, 13000 rpm. The supernatant was discarded by pipetting. The sample was spun briefly

to remove remaining TCA. 1 ml of 90% cold acetone was added, inverted several times and incubated for 15 minutes on ice followed by centrifugation at 4°C, 13000 rpm for 15 minutes. The supernatant was discarded by pipetting and not losing any pellet. 100% acetone (1 ml) was added and the pellet was dried in a desiccator for 15 minutes. The pellet was resuspended in 30µl of SDS-PAGE loading buffer, boiled for 5 minutes at 100°C and 20µl of the sample was applied on a polyacrylamide gel.

2.9.6 Protein identification by Mass spectrometry

Subunits from pACYCD1_MA_S123 were purified from outer membrane using a large batch surface extraction procedure as discussed above. Extracted subunits were boiled at >95°C for 10 minutes, run on Mini-protean TGX pre-cast gel(BioRad) at 200V, stained for 1 hour with Coomassie blue and destain (30% methanol, 10% glacial acetic acid, 60% dH₂O) with multiple changes until no background observed. The proteins from the gel were excised, cut into small pieces and put in a 0.5ml microcentrifuge tube filled with nH₂O. The gels were then destained with 200µl 50% acetonitrile (MeCN) (Rathburn, UK) /50% 10mM triethylammonium bicarbonate (TEAB) (Sigma, UK) overnight. The next day, the gels were dehydrated by adding 200µl MeCN. Then, the MeCN was removed with a SpeedVac (ThermoSavant) for 10 minutes. The gels were hydrated with 200µl TEAB followed by dehydration using the SpeedVac for 10 minutes. It was dehydrated again with 200µl MeCN and removed by SpeedVac for 10 minutes. The gel pieces were reduced with 200µl dithiothreitol (DTT) (Sigma, UK) in 10mM TEAB for 30 minutes at 50°C. DTT was removed and 200µl of 50mM iodoacetamide (IAA) in 10mM TEAB was added and incubated for 30 minutes at room temperature in the dark to alkylate the gels. The IAA was removed, and 10mM

of DTT was added prior to washing (in the dark). This helps reduce N-terminal peptide alkylation. DTT was removed, and gel pieces were washed for three times with 200µl 10mM TEAB and once with 200µl MeCN. Then, gel pieces were dehydrated in SpeedVac for 10 minutes followed by washing with MeCN and TEAB. A total of 200ng of Porcine trypsin (Promega) was added to the dehydrated gel pieces and left to re-swell on ice for 20 minutes. 30 to 50ul of 10mM TEAB was added to fully cover the gel and incubate overnight at 37°C. The next day, the digests were placed on dry ice for 5 minutes, allowed to thaw and transferred to new 0.2ml PCR tubes (Bioquote, UK). A total of 30µl of 10% ACN/5% formic acid (BDH, UK) was added to the gels and placed in a sonicator (Jencons, UK) for 15 minutes and was repeated twice. 30µl of HPLC grade water (Rathburn, UK) was added the samples and analysed by mass spectrometer by staff (Nicholas Micheal) at the Chemical Analysis Facility (CAF), University of Reading.

2.9.7 Purification of *E. coli* SE11 subunits

2.9.7.1 Ni²⁺ affinity chromatography

His6-tagged of *E. coli* SE11 chaperone-subunits bound was prepared in a large batch of 400ml culture of pTRC99_Mhis_S1/DH5α and pTRC99_M8his_S123/DH5α according to periplasmic extraction procedure. HisTrap HP 1 ml (GE healthcare) was fitted on an AKTA purifier system. The AKTA system was washed with degassed nH₂O followed by washing of column with 5 CV (column volume) of elution buffer (20mM sodium phosphate, 500mM NaCl, 500mM imidazole, pH 7.4) with flow rate of 1 ml/min. Then, the column was equilibrated with 5 CV binding buffer (20mM sodium phosphate,

500mM NaCl, 20mM imidazole, pH 7.4) with flow rate of 1ml/min. Sample was applied with flow rate of 0.5ml/min followed by washing with 10 CV binding buffer. The wash fraction and flow-through (FT) fraction were collected. The column bound chaperone-subunits were eluted with 5 CV, flow rate of 0.5ml/min, collecting 1ml fraction in 1.5ml microcentrifuge tubes each. Eluted fractions were kept on ice and analysed by SDS-PAGE.

2.9.7.2 Size exclusion chromatography

Surface extracted subunit 1 and -2 polymer of pACYCD1-MA-S123/BL21 (DE3) constructs was prepared in large batch of 400ml culture. Culture was grown overnight at 26°C in 40ml LB with 0.6% glucose with appropriate antibiotic. The next day, the cells were washed and subcultured into 400ml LB with appropriate antibiotic and grown for 1 hour at 37°C. Then, the culture was induced with 0.5 mM IPTG and grown for another 4 hours at 37°C. After 4 hours, the OD was read. All cells were recovered with Sorvall high-speed centrifuge at 13000 rpm for 30 minutes, 4°C. The recovered cells were resuspend in PBS of 4.5OD/100µl and heat extracted at 55°C. Then the extracted protein was treated with TX114. HiPrep Sepharyl HR S-300 column was fitted on an AKTA purifier system. The AKTA system was washed with degassed 20% ETOH for 10 minutes with flowrate 1ml/min followed by washing of column with 5 CV of nH₂O at 0.5ml/min. Then the column was equilibrated with 2 CV PBS + 0.01% azide at 1ml/min. The proteins were separated with 1.5 CV of PBS at 0.5ml/min and 2 ml samples fraction were collected. The sample fractions were analysed by SDS-PAGE to identify the purified *E. coli* SE11 subunits.

2.10 Bioinformatics analysis

2.10.1 Basic Local Alignment Similarity Search (BLAST)

Sequence similarity searches to identify closely related ushers were performed using BLAST. Usher from *Y. pestis*, (WP_012564968) was used as a query. In this study, BLASTP was used for similarity search of the protein queries against the non-redundant database. Default options were used except for the maximum target sequences, where the number was increased to 1000. All of the sequences were downloaded in FASTA format for further analysis.

2.10.2 CLUSTALW 2.0

ClustalW version 2.0 (Larkin et al., 2007) was used to produce multiple sequence alignments of divergent sequences. This program will align sequences based on the best match. Therefore, the identity, similarity and differences between sequences can be identified. The program is available online from the EBI web server: <http://www.ebi.ac.uk/tools/clustalw2>

2.10.3 PHYLIP 3.695

A phylogenetic tree was constructed using PHYLIP version 3.695 package (Felsenstein, 1993). There are two trees constructed, which are an NJ tree and a bootstrap tree. NJ tree is the original tree estimated using the program and used to interpret the data. However, the bootstrap tree was used to evaluate the NJ tree. The bootstrap tree was constructed by replicating the data set (1,000 replicates for this study) and finally generating a bootstrap tree from all 1,000 trees generated in this process.

Initially, ClustalW2.0 was used to generate multiple sequence alignment (MSA) and the output was saved in *.phy format for used in PHYLIP. For constructing NJ tree, PROTDIST and NEIGHBOR programs were used. The PROTDIST program computes a distance measure for the protein sequences. Then, the distance measure from PROTDIST was used in a distance matrix program, NEIGHBOR. The NEIGHBOR program produced an unrooted tree using the distance matrix method. The NJ was used because this method is very fast and can handle a large data set.

To construct a bootstrap tree a thousand datasets were first made using the SEQBOOT program. Then, the PROTDIST program was used to compute the distance of protein sequences in each dataset. This was followed by producing 1000 trees using the NEIGHBOR program. Finally, CONSENSE program was used to compute consensus tree by the majority-rule consensus method. The bootstrap tree will give values in each node. Node with more than 70% indicates the clade descending from that node is reliable. Subsequently, MEGA5 was used to view the result of the phylogenetic tree.

2.10.4 SignalP

SignalP is a program to predict signal peptide presence and location that discriminate from transmembrane regions. There is option of different organism group (Gram-positive prokaryotes, Gram-negative prokaryotes, and eukaryotes) can be choose in the option. The program can be accessed at <http://www.cbs.dtu.dk/services/SignalP/>

2.10.5 BetaCavityWeb

BetaCavityWeb is a webserver to predict molecular cavities which are important to determine molecular function. This webserver used protein structure to compute the cavities and reports their geometrical properties such as boundary area and narrowest cavity space. The server can be accessed at <http://voronoi.hanyang.ac.kr/betacavityweb>

2.10.6 MEGA 5.2

MEGA 5.2 (Tamura et al., 2011) is a user-friendly integrated program that can help the user to conduct multiple sequence alignments, mining data using the online database and infer phylogenetic trees. In this study, the MEGA program was used to display the output (newick format) from PHYLIP and to analyse and label the tree.

2.10.7 I-TASSER

I-TASSER (Roy et al., 2010) server was used to predict protein structure. In this study, the protein sequence was submitted to I-TASSER and a specific template was assigned. To estimate the quality of the protein model, confidence score (C-score) was used. C-score typically ranges between -5 and 2. Models with a C-score more than -1.5 are expected to have a correct fold. The server can be accessed at <http://zhanglab.ccmb.med.umich.edu/I-TASSER/>

2.10.8 IntFOLD2

InFOLD version 2.0 (Roche et al., 2011) is a server used to predict protein structure and function. It is an independent server that integrates a few programs. The

outputs of the server are tertiary structure prediction (nFOLD), 3D model quality assessment (ModFOLD), identification of disordered region (DISOclust), domain prediction (DomFOLD) and protein-ligand binding residue prediction (FunFOLD). The IntFOLD server is available at the following website: <http://www.reading.ac.uk/bioinf/IntFOLD/>.

2.10.9 TMalign

TMalign program (Zhang and Skolnick, 2005) was used for comparison and alignment of protein structures. This program used dynamic programming iterations to generate residue to residue alignment based on structural similarity. In this study, TMalign was used to obtain model coordinates relative to the template for assembly of the translocon. TM-score was used to assess the reliability of the structure alignment where TM-score between 0.5 and 1 indicates same fold of structural similarity and less than 0.3 indicates random structural similarity. The TMalign program is available at the following website: <http://zhanglab.ccmb.med.umich.edu/TM-align/>

2.10.10 PyMOL version 1.3

PyMOL is an open-source program for protein visualization. This program can produce high-quality 3D images of protein structure and macromolecules for assemblies. At the same time, PyMOL has many options available for protein manipulation. In this study, PyMOL was used in the visualisation of protein structure, model and assembly of Caf proteins. Measurements of the Caf1A barrel were made by using the distance measurement option.

2.10.11 DiscoTope 2.0

DiscoTope 2.0 is a structure-based antibody prediction tool from Immune Epitope Database (IEDB). It was used to predict-cells epitope in Caf1. The DiscoTope 2.0 tool is available at the following website:
http://tools.immuneepitope.org/tools/bcell/iedb_input

CHAPTER 3 ANALYSIS OF F1 EXPORT IN NEW SYSTEM – pACYCF1

3.1 Introduction

In previous studies, the pFMA1 plasmid (Figure 3.1) has been used to study the role of chaperone and usher in the assembly of F1. This pTRC99-based plasmid expresses the three *caf* genes (*caf1M*, *caf1A*, *caf1*), in the absence of Caf1R, from the LacI repressed *P_{trc}* promoter. The plasmid (pFM1) was initially designed for overexpression of the chaperone:subunit complex for structural studies (Chapman et al., 1999; Zavialov et al., 2003). The usher encoding *caf1A* gene was then amplified from pFS2, a construct containing the *caf* operon from *Y. pestis* pFra (Galyov et al., 1990) and cloned in between *caf1M* and *caf1* to form the pFMA1 construct (Chapman et al., 1999). Hence, expression from pFMA1 produces surface assembled F1 but also overexpression of both Caf1M and Caf1A. This system has proved ideal for studying many aspects of chaperone and usher structure and function, as it efficiently traps any nonexported, Caf1 in periplasmic Caf1M:Caf1_n complexes (Dubnovitsky et al., 2010; MacIntyre et al., 2001; Yu et al., 2009; Yu et al., 2012; Zavialov et al., 2002). However, overexpression of usher, in particular, leads to toxicity after 2-3 h induction (Kersley, 2002). Hence high levels of surface F1 cannot be obtained with pFMA1.

To study efficiency of translocation mediated by the Caf1A usher, it is important to use a system where the relative levels of expression of chaperone, usher and subunit closely approximate levels of expression from the native Caf system. One problem with using the subcloned native *caf* operon on a high copy number plasmid, including pFS2 (Galyov et al., 1990), was instability, due to the very high levels of expression of F1. This should be improved with use of a low to intermediate copy number plasmid and controlled expression of F1. Thus, pACYCF1 (R1122) (Figure 3.1) was designed with

the whole native *caf* operon amplified from pFS2 and cloned between the NcoI and SacI sites of pACYCDuet1 plasmid. In this plasmid (pACYCF1), expression of *cafIM*, *cafIA* and *cafI* is under control of *cafIR* regulator.

Surface assembly of F1 from this plasmid does not lead to toxicity therefore higher levels of F1 can be obtained. This Chapter deals with induction of expression and quantitation of F1 expressed from this plasmid.

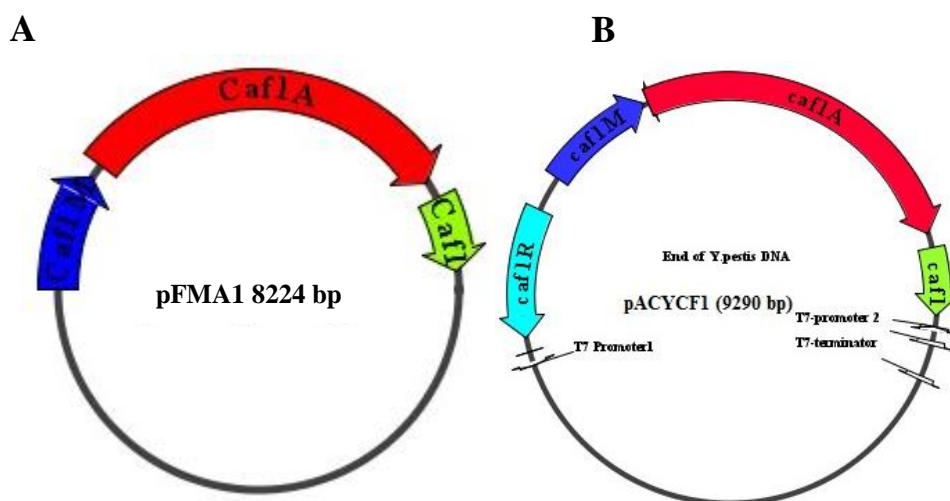


Figure 3.1: Comparison of pFMA1 and pACYCF1 plasmid.

(A) Plasmid used in previous studies (pFMA1), where gene encoding Caf system were individually subcloned into pTRC99 plasmid, produces a low level of CafI and overexpression of *cafIM* and to lesser extent *cafIA*, under control of *ptrc* promoter (IPTG induction). (B) Plasmid used in this study (pACYCF1) where the native *caf* operon cloned in pACYCDuet1, expression of *caf* genes under control of *cafIR*.

3.2 Comparison of assembled F1 and chaperone from pACYCF1 and pFMA1

This study began by comparing levels of CafIM and F1 produced from pACYCF1 and pFMA1. Both plasmids were transformed into competent *E.coli* DH5 α

cells and one transformant from each was induced with 0.5 mM IPTG (pFMA1) for 1.5 h or thermo-induced (pACYCF1) for 4 hours as in method 2.8. Chaperone denatures at 55°C (Zavialov et al., 2005); hence F1 can be readily extracted from the cell surface by heating induced cells at 56°C for 1 hour. Both whole cells and surface extracted F1 were analysed by SDS-PAGE. F1 is very stable and only denatures at 75°C in SDS or at 90°C without SDS. Therefore, samples were applied to SDS-PAGE unheated to detect Caf1 polymer (F1) or following heating at >95°C for 15 minutes, in SDS-PAGE sample buffer to visualise and quantitate denatured Caf1 subunit. Figure 3.2 shows that the level of F1 in whole cells and surface extracted F1 from *E. coli* DH5 α /pACYCF1 is higher compared to that from cells following induction from pFMA1. This is despite the high level of chaperone visible in samples from *E.coli*/pFMA1 (see Figure 3.2).

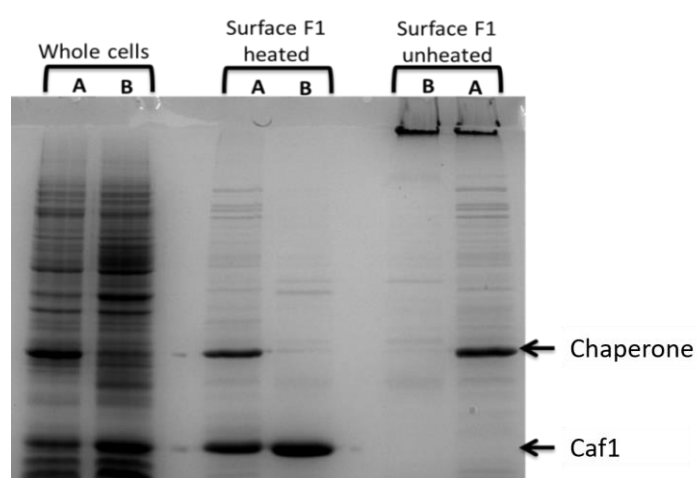


Figure 3.2: SDS-PAGE analysis to compare level of surface F1 in *E. coli* DH5 α carrying either pACYCF1 or pFMA1.

Following induction for 1.5 hours with IPTG for pFMA1 (A) and 4 hours at 37°C for pACYCF1 (B), 3ODs of cells were recovered and heat shocked in 100 μ l PBS to isolate surface F1. Whole cells were resuspended at 1OD/200 μ l cells. 10 μ l of each sample was loaded in well.

Induced cells of *E. coli* DH5 α /pACYCF1 and *E. coli* DH5 α /pFMA1 were also compared for the level of periplasmic Caf1M. Periplasmic fractions from both were isolated by osmotic shock and analysed by SDS-PAGE. Figure 3.3 highlights the much lower level of expression of Caf1M (26.3 kDa) from *E. coli* DH5 α /pACYCF1 compared to pFMA1. Identity of Caf1M was confirmed by immunoblotting (Figure 3.3

(B)). In general, therefore, it seems that despite the low level of chaperone in pACYCF1, assembly of F1 on the cell surface is more efficient than from pFMA1. With the very low level of chaperone produced for pACYCF1, it is likely that non-exported Caf1 would not accumulate in the periplasm but be degraded.

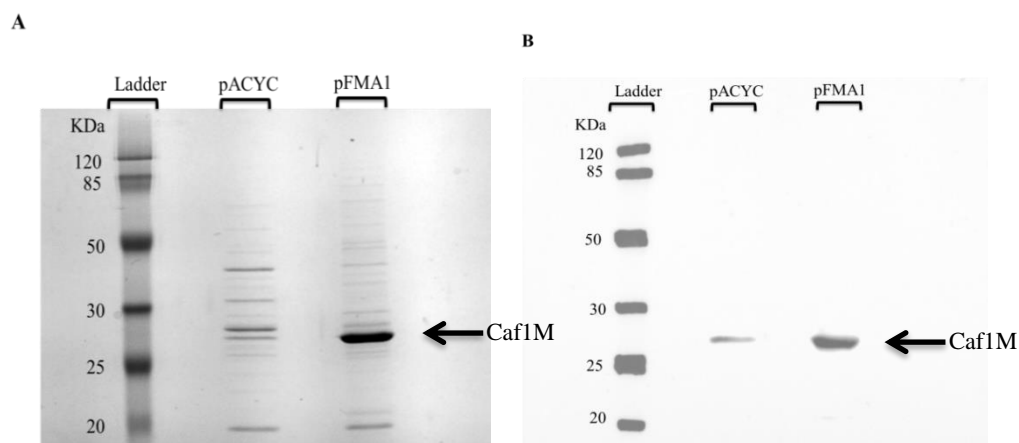


Figure 3.3: Comparison of level of Caf1M chaperone in induced cultures of *E. coli* DH5α carrying pACYCF1 or pFMA1. (A) SDS-PAGE of periplasmic fractions stained with Coomassie Blue. (B) Immunoblotting of periplasmic fraction with 1 in 10000 dilution of rabbit anti-Caf1M serum. 10μl sample was loaded in each lane.

3.3 Optimisation of induction time and Caf1 extraction temperature

3.3.1 Thermoinduction of pACYF1

To test the induction time required for expression of Caf1, pACYCF1 (R122b) was retransformed in *E. coli* DH5α. Six individual colonies were cultured ON at 26°C then induced for 4 hours or 6 hours at 37°C. Cells from each culture were resuspended at 4 OD equivalent/ml and analysed by SDS-PAGE for F1 production. Figure 3.4 shows samples of heat denatured F1 in whole cells. Caf1 is clearly visible and is the most prominent band in all samples. Despite the fact that samples induced for 6 hours have

more Caf1 compared to samples induced at 4 hours, 4 hours was selected as the best time to monitor differences. This is because at 4 hours, the cells are still growing (exponential phase).

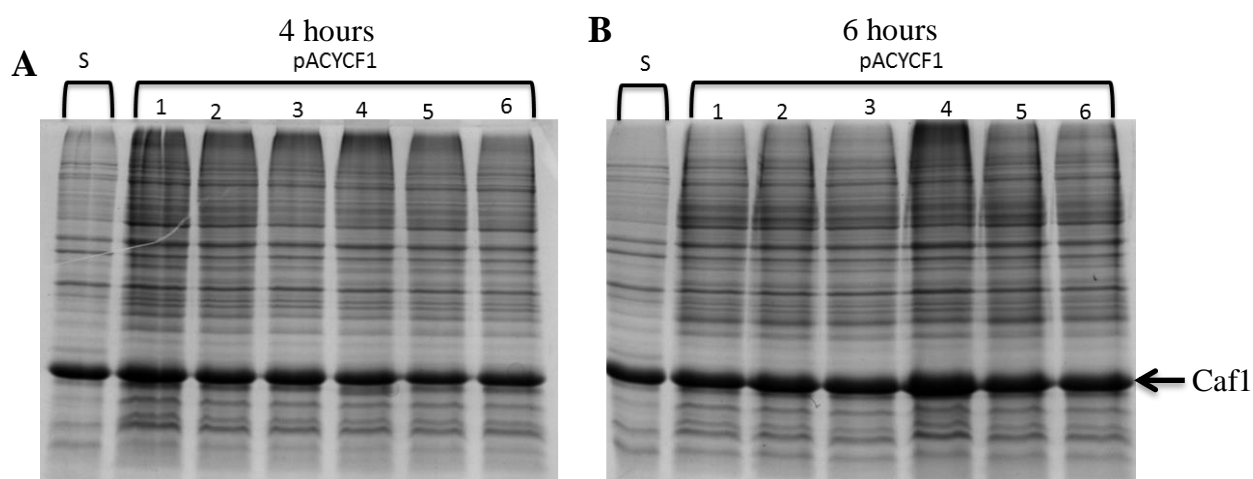


Figure 3.4: Analysis of F1 production in six transformants of *E. coli* DH5 α /pACYCF1 expression from whole cells.

(A) Production of F1 at 4 hours induction. (B) Production of F1 at 6 hours induction. 1 to 6 indicate different transformants.

To establish the optimal induction temperature of F1 from pACYCF1 individual colonies were cultured ON at 26°C then subcultured and induced at 26°C, 28°C or 37°C. F1 was analysed directly in heat-denatured whole cells and also as surface extracted protein by SDS-PAGE (Figure 3.5). In Figure 3.5 (A), there is no Caf1 evident at 26°C while it is clearly visible at 28°C. However, in Figure 3.5 (B), very low levels of F1 were detected following surface extraction of cells cultured at 26°C. The level of F1 in whole cells and surface extract of cells induced at 28°C were slightly less than that in cells induced at 37°C. Thus, a temperature of 37°C and induction time of 4 hours was used for subsequent experiments in this study.

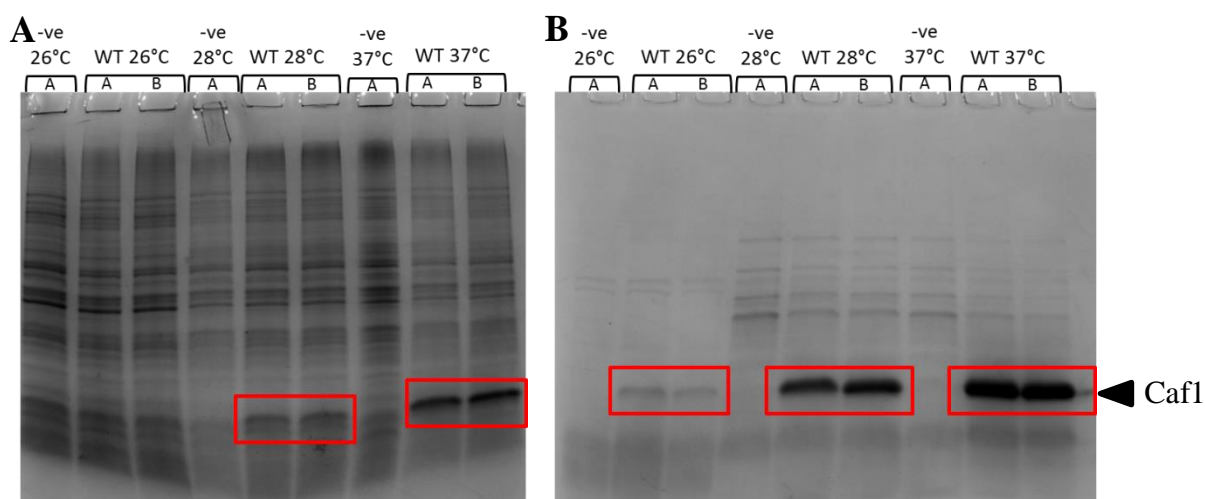


Figure 3.5: SDS-PAGE analysis of thermoregulation of F1 at 26°C, 28°C and 37°C. (A) Whole cells of *E. coli* DH5α/pACYCF1 (WT) following 4 hours subculture at indicated temperature. (B) Surface extracted F1, red box indicates heat denatured F1. -ve indicates, *E. coli* DH5α/pACYCduet1.

3.3.2 Optimisation of F1 heat extraction for quantitation

Extraction of F1 from whole cells facilitates quantitation. This was routinely performed between 55°C-57°C for 1 hour. However, with large amounts of F1, all F1 was not always completely extracted. Therefore, prior to quantitation, the efficiency of extraction of F1 at 65°C was compared to that at 57°C. Two transformants from the *E. coli* DH5α/pACYCF1 were inoculated ON and thermo-induced at 37°C for 4 hours. 3 ODs of cells were recovered by centrifugation at 13000 rpm for 20 minutes, then heat shocked at 57°C and 65°C for 1 hour subsequently to extract surface F1. Extracted F1 was analysed by SDS-PAGE (Figure 3.6). The level of extracted F1 is higher from cells extracted at 65°C compared to 57°C heat shocked samples. It is also clear that there is a higher polymer level in the stacking gel of 65°C heat shocked samples compared to 57°C (Figure 3.6 (B)). SDS-analysis of pelleted cells after surface extraction of F1

shows that a substantial F1 band remained associated with whole cells after heating at 57°C, whereas with the 65°C sample this was greatly reduced (Figure 3.6 (C)). In conclusion, heat extraction of F1 at 65°C was very efficient and recovered most surface F1. Therefore, 65°C was used to isolate surface F1 for quantitation from cells expressing loop mutation constructs (refer Chapter 5).

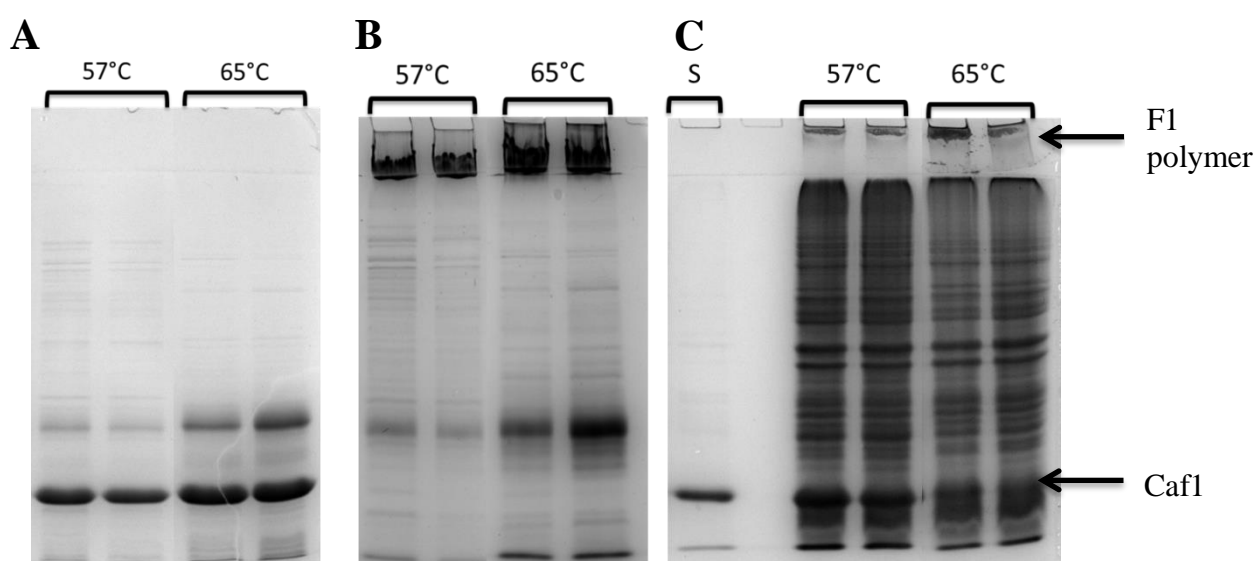


Figure 3.6: Impact of extraction temperature on recovery of surface F1. F1 polymer was extracted from cells of *E. coli* DH5 α /pACYCF1, thermoinduced for 4 hour, by incubation at 57°C or 65°C for 1 hour

(A) Heat denatured, extracted F1. (B) Unheated extracted F1 polymer. (C) Heat denatured whole cells pelleted after surface extraction. For denaturing, 18 μ l of sample was mixed with 6 μ l of 4X loading buffer and heated at 99°C for 15 minutes. 20 μ l was loaded onto the well. S indicates F1 standard.

3.3.3 Comparison of Caf1 from whole cells, surface extracted and TCA

precipitation of culture media

TCA precipitation (see section 2.9.6 for method) was used to assess the amount of F1 that was released into the culture media from the cell surface of *E. coli*

DH5 α /pACYCF1 during thermoinduction for 4 hours. Two transformants were inoculated, grown ON (26°C) and thermos-induced. The cells were recovered by centrifugation at 13000rpm for 20 minutes for whole cell analysis (4 OD/ml), surface extraction of F1 (3 OD/100ul) and TCA precipitation of F1 in spent growth media (~1 OD/30ul). Samples were heated for 15 minutes at >95°C prior to SDS-PAGE. Figure 3.7 shows a faint band of F1 that had remained in the supernatant after 20 minutes centrifugation. However, the amount of F1 released from the cell surface into the media during the 4 hours growth was very low, in comparison to that recovered by extraction from the cell surface. Therefore, it was concluded that surface extraction was the most effective method for efficient recovery and quantitation of F1.

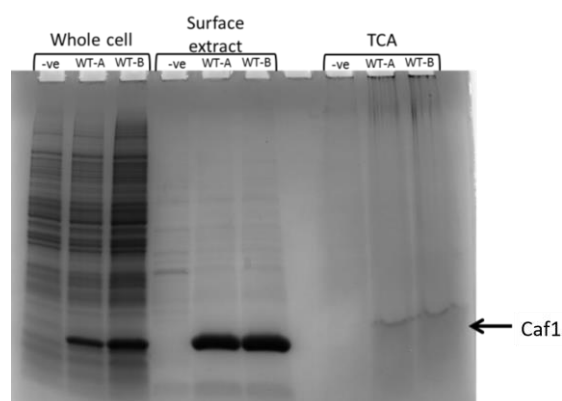


Figure 3.7: SDS-PAGE analysis of F1 in whole cells, surface extraction and TCA precipitate from culture media.

All samples were heated at >95°C for 15 minutes. 10 μ l of whole cells and surface extracted F1 were applied to the gel while 20 μ l of the TCA precipitate from spent culture media was applied. The level of Caf1 in WT-B of whole cells extracted was higher because the amount of cells applied was more compared to WT-A.

3.4 Caf1 quantitation following SDS-PAGE

As accumulation of F1 in the periplasm cannot be used to monitor efficiency of surface assembly with expression from pACYCF1, a reliable method of quantitation of

surface F1 was required. Because of the high levels of Caf1 visible in both whole cells and surface extracts, quantitation by imaging denatured Caf1 resolved by SDS-PAGE offered a direct quantitative assay. Initially, a large batch of F1 was prepared by heat extraction, as standard. This preparation contained primarily F1 but also a low level of cell contaminants (Figure 3.8). The total protein concentration in the original F1 standard was calculated as 345.03µg/ml using the Bradford assay and gamma globulin as standard (Table 3.1). The F1 concentration was calculated by imaging and quantitation following SDS- PAGE and average percentage of F1 band on each lane was 85%. Therefore, the concentration of F1 from each lane was estimated at 293.25µg/ml.

Table 3.1: Estimated concentration of total protein in F1 standard.

Duplicate samples (4-8 µl) were made up to 150µl with PBS for the Bradford microassay. Protein concentration in this 150µl was calculated from the standard curve. The dilution factor for preparation of the 150µl test sample was then used to calculate the estimatec total protein concentration in the original F1 standard preparation.

Sample Applied (µl)	OD595	OD595	Mean OD595	Standard deviation	Concentration (µg/ml)	Dilution	Original concentration (µg/ml)
8	0.28	0.265	0.2725	0.011	23	18.75	430.1
6	0.209	0.189	0.199	0.014	14	25	350
4	0.11	0.115	0.1125	0.003	6.8	37.5	255
Average							345.03

To accurately compare levels of F1 produced under different conditions or from different mutants, the sensitivity of the imaging and quantitation assay was assessed by quantitating different levels of F1. GeneTools was used to calculate the amount of F1 applied based on intensity of band and subtraction of background (See section 2.9.2). A graph was plotted based on the exact amount of F1 applied against Caf1 band raw

data (gel imaged) to determine the reliability and proportionality of quantitation of the F1 band with increasing sample concentrations.

Figure 3.9 shows that the assays were not proportional. High levels of applied samples give proportionally slightly lower readings for denatured Caf1 than low levels. Therefore, an amount of F1 standard of 2.5-3.75 μl was used to apply on a gel for quantitation because this range is the most linear point in the graph.

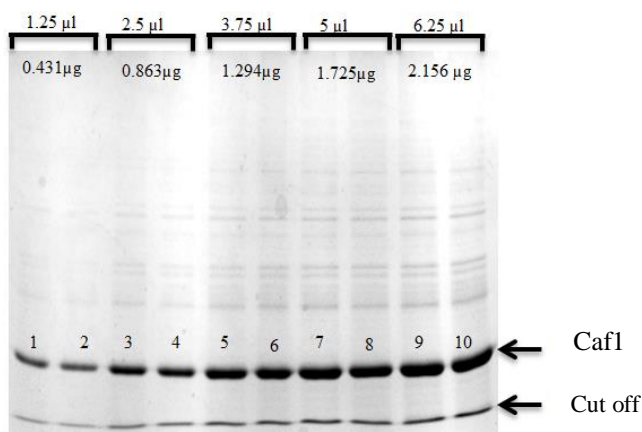


Figure 3.8: Standard gel for F1 quantitation.

Increasing amount of F1 from 2 μl to 10 μl were made up with dH_2O to 12 μl and mixed with 4 μl of 4X sample loading buffer, denatured and 10 μl was applied for SDS-PAGE. μl indicates volume of original F1 standard in each sample applied to the gel and μg indicates the total protein applied, based on the Biorad estimated protein concentration of the F1 preparation (0.345 $\mu\text{g}/\mu\text{l}$).

Table 3.2: The raw data of F1 standard from each lane of SDS PAGE. The percent was calculated from the raw data to get the concentration of Caf1 protein (F1) within the crude F1 preparation standard.

Lane	Exact amount applied (μl)	F1 band raw data	F1 band raw data (average)	Total F1 protein (μg)
1	1.25	4229.13	4277.76	0.431
2	1.25	4326.4		
3	2.5	7074.42	7280.27	0.863
4	2.5	7486.13		
5	3.75	10211.5	9753.4	1.294
6	3.75	9295.32		
7	5	11633.9	11213.9	1.725
8	5	10794.00		
9	6.25	13102	13023.4	2.156
10	6.25	12944.8		

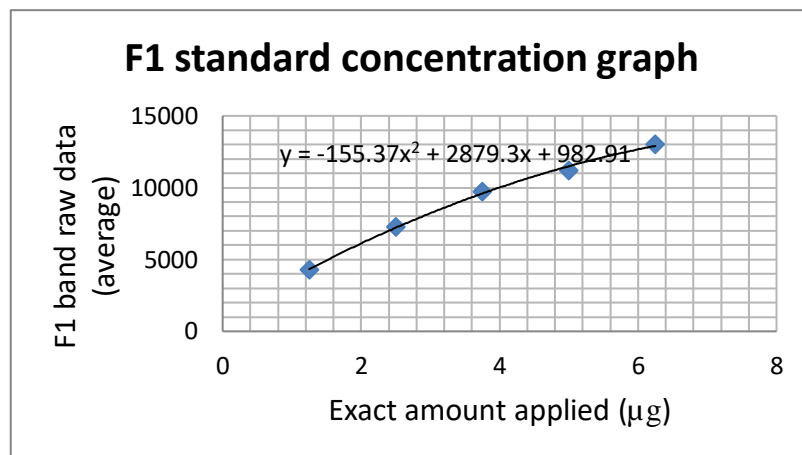


Figure 3.9: F1 standard concentration graph for determination of F1 standard reliability.

3.5 Conclusions

In this chapter, it is clear that the expression of recombinant *cafI* from pACYCF1 in *E. coli* is thermoregulated as has previously been noted from *Y. pestis* (Brubaker, 1972; Perry and Fetherston, 1997) and other recombinant constructs (Chapman et al., 1999; MacIntyre et al., 2001). In this system, production of F1 polymer was ‘on’ above 28°C and ‘off’ below 26°C. Therefore, throughout this study, to ensure stability, *E. coli* cells carrying pACYCF1 were always cultured at 26°C until expression was required and growth was switched to 37°C.

Accumulation of periplasmic CafI has been used as an assay to monitor inefficiency of export across the outer membrane (Zavialov et al., 2002). However, non-exported periplasmic CafI subunit is mainly degraded (Chapman et al., 1999; MacIntyre et al., 2001). Therefore, with pACYCF1 and the low levels of chaperone, it is likely that most non-assembled CafI would be degraded. Hence, accurate direct quantitation of extracted surface F1 is required. Chaperone denatures at 55°C (Zavialov

et al., 2005) and heat extraction of bacterial cells coated in F1 polymer at this temperature effectively releases F1 from the cell surface. Here, it is shown that increasing the extraction temperature to 65°C for 1 hour enhances F1 surface extraction resulting in a very efficient extraction. TCA precipitation of the spent culture media demonstrated that very little F1 was released into the growth medium within the first 4 hours induction. Hence, monitoring F1 levels in whole cells and surface extracts is a valid assay for comparisons of F1 production.

A reliable F1 quantitation assay, based on imaging following SDS-PAGE and comparison to an F1 standard has been established. However, the assay is not directly proportional. Therefore, two levels of standard were routinely used for subsequent quantitation – 2.5µl (0.863µg) for comparison with poorly functioning mutants and 8.5µl (2.59µg) for production closer to wildtype.

CHAPTER 4 MODELLING AND PROPERTIES OF THE CAF TRANSLOCON

4.1 Introduction

This chapter is focused on understanding details of the structure and function of the *Yersinia pestis* translocon of the Caf system. The N-terminal domain of Caf1A had previously been shown to participate as the primary binding site of the Caf1M:Caf1 complex (Phan et al., 2011; Di Yu et al., 2012; Ng et al., 2004). Deletion of Caf1A N-terminal domain results in accumulation of Caf1:Caf1M complexes in the periplasm and completely abolished export of F1 to the cell surface (Di Yu et al., 2012; Yu et al., 2009). The Caf1A middle domain is important in plugging the pore (Remaut et al., 2008; Yu et al., 2009) and Caf1A C-terminal domain is crucial in assembly of Caf1 polymer (Dubnovitsky et al., 2010). Two hydrophobic clusters from Caf1A C-terminal domain (Cluster 1: Phe740, Iso743, and Iso755, and Cluster 2: Phe727, Tyr763, and Ala800) were demonstrated as essential for assembly. Double mutations of these residues resulted in accumulation of Caf1 polymer in the periplasm.

Structures of Caf1A usher N-terminal, C-terminal and middle domains have been solved (refs). Other available structures of the Caf system include, wild type and mutant complexes of Caf1M:Caf1:Caf1 complexes (Zavialov et al., 2003, Zavialov et al., 2005, Yu Di et al 2012), subunit-free Caf1M (Di Yu et al., 2012), Caf1A N-terminal domain in complex with Caf1M:Caf1 as well as alone (Di Yu et al., 2012). The only remaining unsolved structure in the Caf system is the Caf1A usher barrel. To advance our understanding of usher function and facilitate strategic applications of the F1 assembly system, structural information of the Caf1A usher and the Caf translocon is very important. Therefore, the structure of FimD (Phan et al., 2011; Geibel et al., 2013) was used to model open Caf1A and the structure of PapC (Remaut et al., 2008) to model

closed Caf1A using homology modelling. The model of open Caf1A usher was then assembled with the available structure of Caf1:chaperone complex (Zavialov et al., 2003) to create the Caf translocon. Potentially important interacting residues were identified from alignments and location of conserved residues within the modelled translocon. The importance of identified residues for F1 polymer assembled was addressed by mutagenesis. This modelled translocon was also used in Chapter 5 to optimise export of F1 with short peptide inserts.

4.2 Extension of γ 3 family of fimbrial usher protein

A previous publications identified 15 members of the γ 3 family of fimbrial usher proteins with the closest relative of Caf1A usher as *Y. pestis* CO92 (PsaC) and *Y. enterocolitica* (MyfC) (Nuccio and Baumber, 2007). Availability of sequence information of closer relatives might aid in identifying key residues in usher proteins of this family. Therefore, the BLASTP programme from NCBI was used to search the non-redundant database for additional members of γ -3 family of ushers and in particular for close relatives of Caf1A. The number of hits in the search was increased to 1000, identical entries were excluded and results downloaded and summarised in an excel file (Appendix 1). From the similarity results in Appendix 1, there are 34 new sequences that had not already been assigned to the γ fimbrial group (Nuccio and Baumber, 2007). These include usher proteins from *E. coli* SE11, *S. enterica*, *E. coli*, *Klebsiella pneumonia*, *Klebsiella oxytoca*, *Serratia marcescens*, *Serratia odorifera*, *Y. enterocolitica*, *Enterobacter cancerogenus*, *Providencia rustigianii*, *Cronobacter dublinensis*, *Klebsiella Sp.*, *Cronobacter sakzakii*, *Serratia fonticola*, *Edwardsiella piscicida* and *Enterobacter hormaechei*. By compiling the 34 new sequences together

with previously identified proteins in the γ fimbrial group, a phylogenetic tree was constructed to expand the γ group.

Figure 4.1 shows that of the 34 newly identified ushers, the $\gamma 3$ fimbriae group contained eight new putative ushers annotated as follows; AafC from *E. coli* 042, AAD27810 (confirmed with 100% bootstrap), an usher protein from *E. coli* 1-250-04_S1_C3, EZJ94256 (confirmed with 96% bootstrap), a fimbrial usher protein from *E. piscicida*, GAJ67292 (confirmed with 100% bootstrap), an usher protein from *E. coli* SE11, WP_000877656.1 (confirmed with 100% bootstrap) and four putative fimbrial usher proteins from *S. enterica* (WP_001750818, WP_023222330, WP_001750512, WP_023222950) (confirmed with 100% bootstrap value). The whole $\gamma 3$ fimbriae group was confirmed with 87% bootstrap value. This excludes a type I fimbriae anchoring protein FimD (WP_021806158.1) from *Serratia fonticola* which was identified with bootstrap value of only 37%. For this protein the bootstrap value was only 47% which indicates that it is unreliable to include this usher in the $\gamma 3$ fimbriae group. It is clear that the usher from the commensal *E. coli* SE11 grouped with $\gamma 3$ fimbriae. It is also clear that from the phylogenetic tree that the usher from *E. coli* SE11 (Oshima et al., 2008) lies next to *Y. pestis* Caf1A indicating a close evolutionary relationship with 70% identity within 99% sequence coverage. They are likely to share common functional features with respect to assembly via the chaperone-usher system. The new *S. enterica* clade that grouped in $\gamma 3$ also clustered close to Caf1A from *Y. pestis*. This is interesting in view of the recognition that *Y. pestis* pMT1/pFra plasmid shares common ancestry, but not the *caf* locus, with a cryptic plasmid of *S. enterica* (Prentice et al., 2001). This interesting relationship between the *Y. pestis* Caf locus, the

related chaperone-usher loci from *E. coli* SE11 and *S. enterica* loci is investigated in more depth in Chapter 6. Based on the previous publication (Nuccio and Baumler, 2007), the most distant ushers in the $\gamma 3$ fimbriae group are the CS3 and CSSD usher proteins. This was confirmed in this expanded $\gamma 3$ fimbriae group.

A new clade was also identified in the γ fimbriae group (see Figure 4.1, grey clade). This clade was only supported with a bootstrap value by 25%. Therefore, a new phylogenetic tree was constructed along with members of all fimbriae usher family to identify the group to which these usher proteins belong. To construct the phylogenetic tree, three members from each fimbrial usher family were picked randomly, and along with the ushers from the undefined group (grey clade), a phylogenetic tree was constructed using PHYLIP package. Based on Figure 4.2, it is clear that this undefined group belongs to the Gamma fimbriae family, however, this group of ushers form a distinct clade. Thus, further analysis must be done to this clade to characterise the ushers.

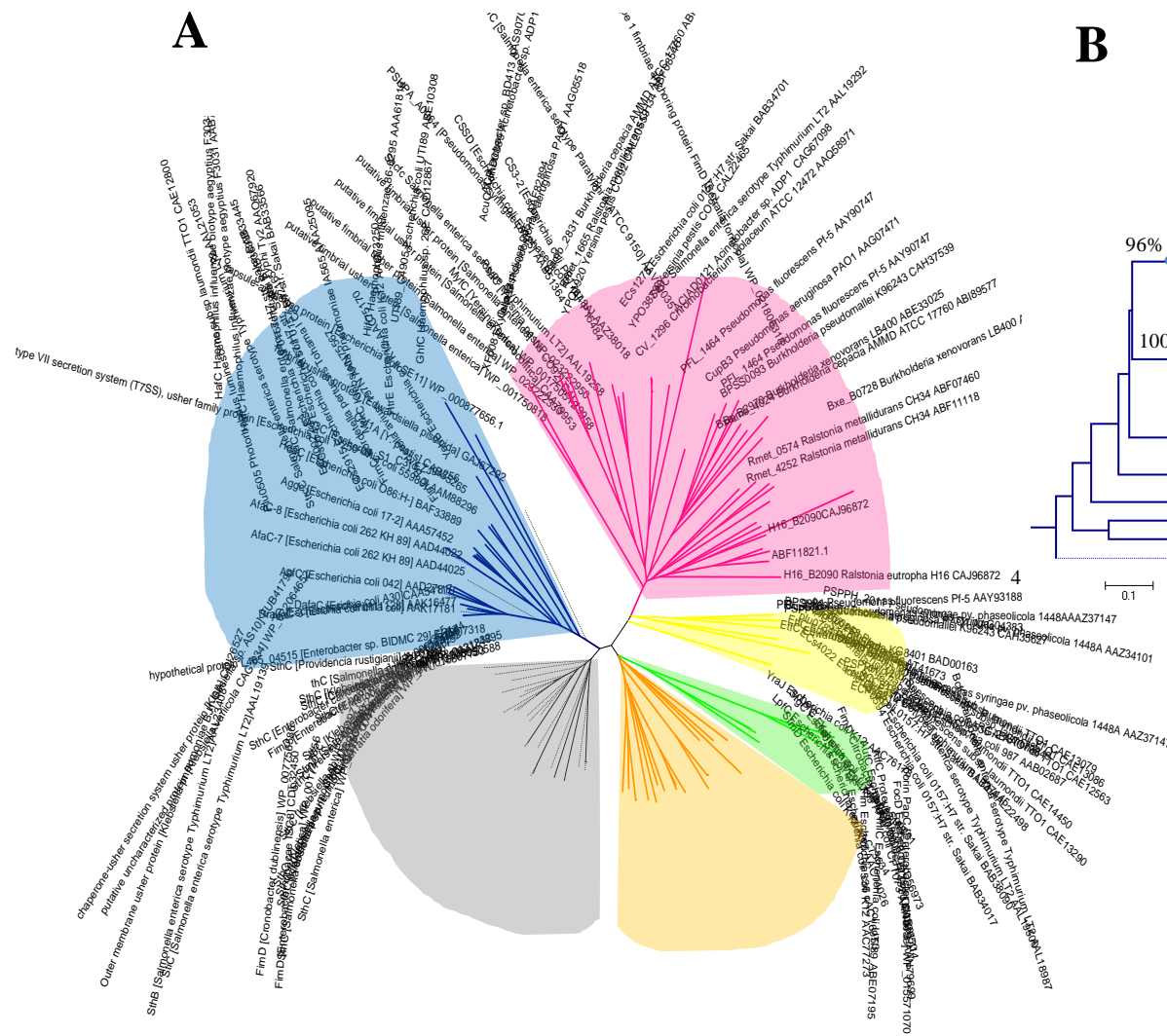


Figure 4.1: Phylogenetic tree of Gamma fimbriae group of usher proteins.

(A) Coloured in yellow is γ , orange is γ_1 , green is γ_2 , blue is γ_3 , pink is γ_4 and grey is undefined group. (B) γ_3 fimbriae subtree. Type 1 fimbriae anchoring protein FimD (*S. fonticola*) was excluded from the γ_3 fimbriae group because the bootstrap value was only 47%. All other proteins in γ_3 were supported with 87% bootstrap value. Pale blue, dotted line indicates newly identified proteins that cluster in the γ_3 usher clade (refer Appendix 1) and blue, linear line indicates proteins previously identified as belonging to γ_3 group (Nuccio and Baumler, 2007)

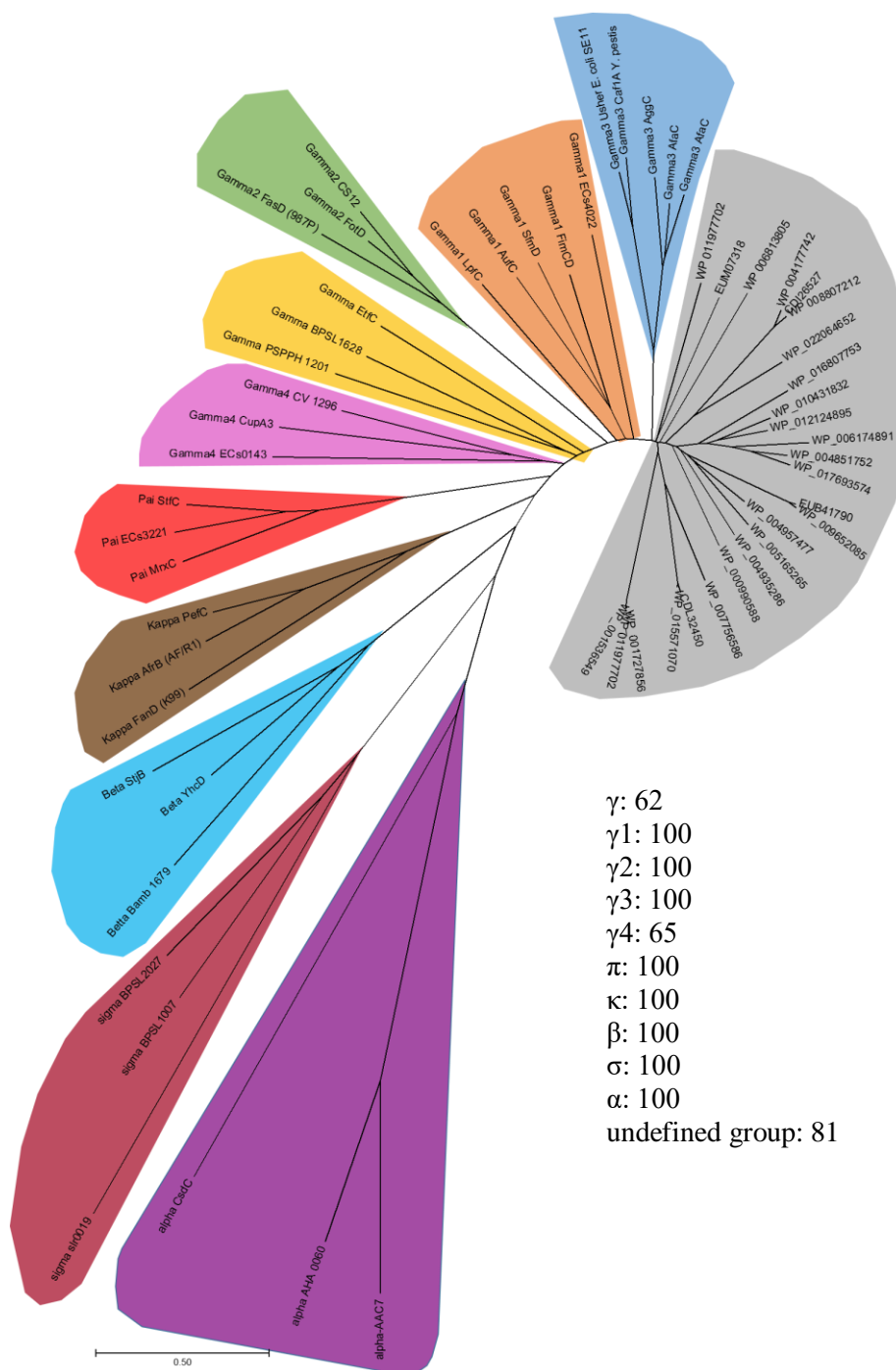


Figure 4.2: Phylogenetic tree of fimbrial usher protein family.

Colored in yellow is γ , orange is $\gamma 1$, green is $\gamma 2$, blue is $\gamma 3$, pink is $\gamma 4$, red is π , brown is κ , cyan is β , maroon is σ , purple is α and grey is undefined group. Number indicates bootstrap percentage for each family of 1000 replicates.

4.2.1 Structural proteins loci of the $\gamma 3$ family members.

Generally, $\gamma 3$ -fimbriae consist of one major subunit that forms the fibre while some members of this group contain a second subunit (Nuccio and Baumler, 2007; Zav'yalov et al., 2010). The $\gamma 3$ fimbriae operons encode a subunit(s) that is assembled into thin, flexible fimbriae, in contrast to the rigid multi-subunit fimbriae of the $\gamma 1$ group, for example Type I fimbriae assembled by FimD usher. The $\gamma 1$ -fimbriae, in addition, include a terminal adhesin, approximately twice the size of subunit proteins. Therefore, to provide detailed consideration of usher function of the extended $\gamma 3$ -fimbriae group, the entire gene cluster was documented, initially using the NCBI database. For some new members of the $\gamma 3$ group, only putative ushers were identified, the remainder of the operon was not annotated. The core structure for a functional operon in this group would be genes encoding at least one subunit, an usher and a chaperone. Some operons encode one regulator for example *caf* system, some two very different regulatory proteins for example *psa*; while others have no associated regulator gene, including the gene cluster from *E. coli* SE11. The operon from *E. coli* SE11 consists only of one chaperone, one usher and two subunits. (See Chapter 6 for detailed analysis). While both chaperone and usher can be readily identified by BLAST analysis, identification of subunits is more difficult and generally awaits experimental confirmation. However, based on size and location in the operon, it is possible to predict genes encoding structural subunits.

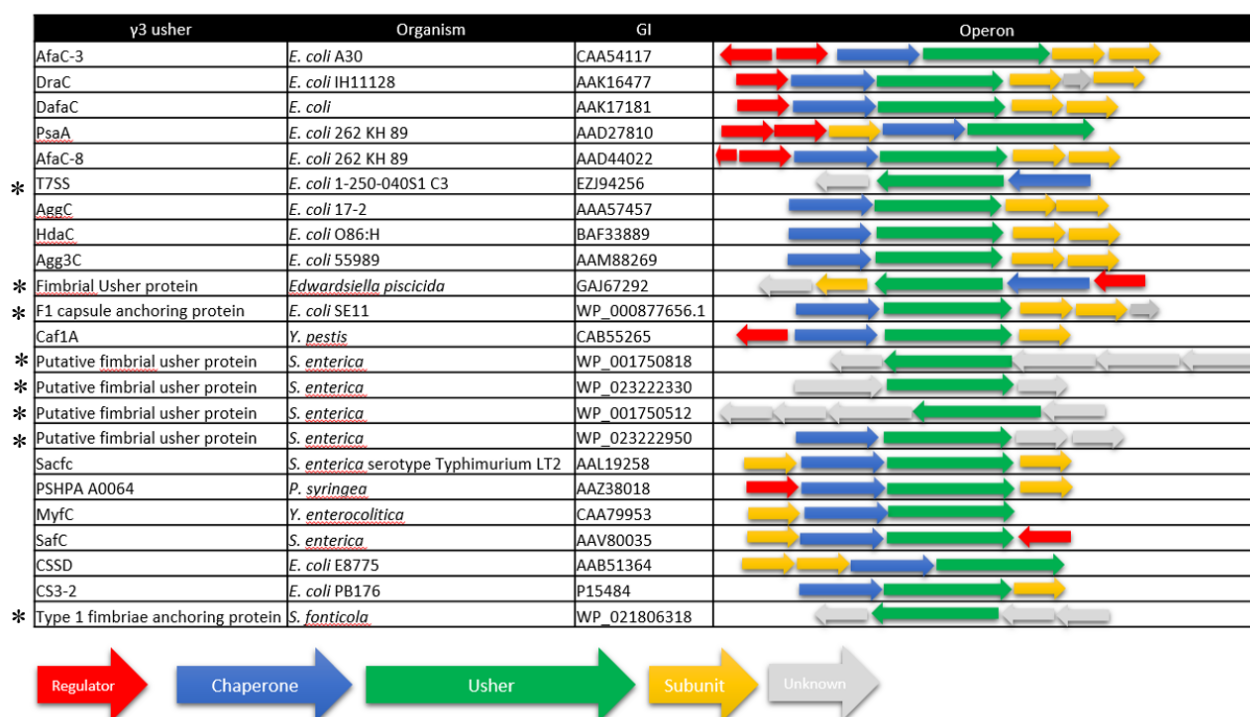


Table 4.1: Locus structure of updated $\gamma 3$ fimbriae group CU systems.

Red arrow represents regulator, blue chaperone, green usher, yellow subunit and grey unknown protein. Protein with asterisk (*) is newly identified usher in $\gamma 3$ fimbriae group.

4.3 Operon analysis and domain analysis of *E. coli* SE11 caf like genes

The most similar protein to Caf1A was an usher from *E. coli* commensal SE11 (accession no WP_000877656.1) (Oshima et al., 2008) with an E-value of 0 and 70% identity. This gene lies on plasmid pSE11-2 DNA (NC_011413.1). This is remarkably close to Caf1A; the next closest was *E. coli* AfaC, found to have an E-value of 0 and 44% identity. Because of the very close identity of this *E. coli* SE11 usher to *Y. pestis* Caf1A this chaperone usher locus was studied in more detail. The context of this usher gene within the *E. coli* SE11 plasmid as recorded in the Entrez database is shown in Figure 4.3. The usher from this bacterium was clustered with a chaperone and three putative structural subunits.

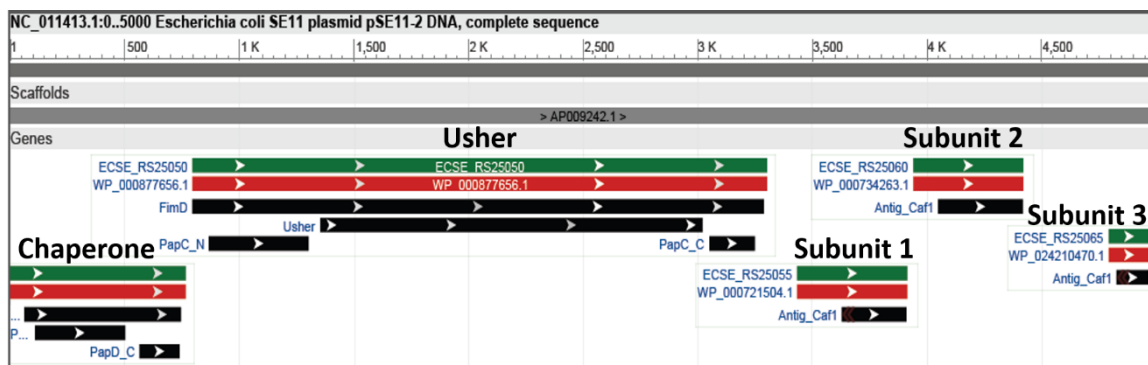


Figure 4.3: *E. coli* SE11 plasmid encoded chaperone-usher cluster (pSE11-2 (NC_011413.1)).

The upstream and downstream sequences next to this cluster encode hypothetical proteins and do not identify any domain when checked for domain in Interpro. The data was obtained from Entrez database.

Protein analysis of proteins within this SE11 CU cluster was performed by using Interpro to confirm domain identification and BLASTP to look for sequence similarity against the non-redundant database. As expected and as shown in Table 4.2, *E. coli* SE11 chaperone had a PapD-like domain predicted from CATH and Superfamily database, PapD-N domain predicted from Pfam and NCBI CDD (Conserve Domain Database), and PapD-C domain predicted from CATH, Superfamily and Pfam database. Caf1M is also predicted to have these domains. As for similarity search, the closest relative to the chaperone from *E. coli* SE11 was Caf1M from *Y. pestis* with an E-value of 2E-109 and 59% identity.

As expected from the annotation, *E. coli* SE11 usher from this locus (WP_000877656.1) exhibits high similarity to Caf1A usher protein from *Y. pestis* with an E-value of 0 and 70% identity. Domain search predicted the same domain structure as Caf1A with PapC_N, usher and PapC_C from both NCBI CDD and Pfam database. The first subunit (WP_000721504.1) was predicted to have an antigen_Caf1 domain and similar

to Caf1 from *Y. pestis* with an E-value of 4E-04 and identity of 34% over the whole protein. The second (WP_000734263.1) also shared similarity with Caf1 with identity of 31% and E-value of 6E-04. Interestingly, subunit 2 was also predicted to have a Sef14_adhesin domain (Pfam). Neither this subunit nor Sef14 have the two domain structure and size of the terminal adhesins as seen in Pap and Fim pili. Hence, these can be predicted to have the single Ig-domain pilin fold. Due to the small size of the third gene ((WP_024210470.1, only 58 amino acids), it would not be capable of forming a typical pilin-fold and is unlikely to be important as a structural component of this operon (see Table 4.2). Chapter 6 discusses of *E. coli* SE11 plasmid 2 (pSE11-2) encoded chaperone-usher system in more detail.

Table 4.2: In silico analysis of *E. coli* SE11 pSE11-2 chaperone-usher cluster. Pairwise alignments were performed against mature Caf1 and mature *E.coli* SE11 subunits by EMBOSS Needle programme.

<i>E. coli</i> SE11 CU cluster proteins	Domain analysis (Interpro)	Similarity search (BLASTP)
Chaperone (254 aa) (WP_012564967.1)	PapD_N (NCBI CDD) PapD-like (CATH, Superfamily) PapD-N (Pfam) PapD-C (CATH, Superfamily and Pfam)	F1 chaperone protein, <i>Y. pestis</i> (WP_002211762.1) E-value: 2E-109 Identity: 59% Query cover: 100%
Usher (835 aa) (WP_000877656.1)	Usher (Pfam, Panther, NCBI CDD) PapC_N (Pfam, NCBI CDD) PapC_C (Pfam, NCBI CDD)	Fimbrial usher protein, <i>Y. pestis</i> (WP_002211763.1) E-value: 0 Identity: 70% Query cover: 99%
Subunit 1 (159 aa) (WP_000721504.1)	Capsule antigen, Caf1 (NCBI CDD , CATH and Pfam)	Capsule protein F1 (precursor), <i>Y. pestis</i> (AGX27780.1) E-value: 5E-04 Identity: 34% Query cover: 99%
		Pairwise alignment of mature subunits Identity: 30.4% Similarity: 45.6%

Subunit 2 (159 aa) (WP_000734263.1)	Capsule antigen, Caf1 (NCBI CDD , CATH and Pfam) Sef14_adhesin (Pfam)	Capsule protein F1 (precursor), <i>Y. pestis</i> (AGX27780.1) E-value: 6E-04 Identity: 31% Query cover: 100%
		Pairwise alignment of mature subunits Identity: 29.1% Similarity: 48.3%
Subunit 3 (58aa) (WP_024210470.1)	Capsule antigen, Caf1 (NCBI CDD , CATH and Pfam)	Capsule protein F1 (precursor), <i>Y. pestis</i> (AGX27780.1) E-value: None Identity: None Query cover: None
		Pairwise alignment of mature subunits Identity: 14.1% Similarity: 22.8%

4.4 Modelling of the Caf translocon

Following review of available sequences of the $\gamma 3$ fimbrial usher protein family, protein modelling of Caf1A was performed. Availability of related sequences should aid in determining key residues that are involved in the function and assembly of F1. Available crystal structures of the Caf system include, wildtype (PDB:1Z9S, 2.2 Å) and mutant (PDB:1P5U, 1.99 Å) of Caf1M:Caf1:Caf1 complexes, (Zavialov et al., 2005), the Caf1A C-terminal domain (PDB:2XET, 1.6 Å) (Dubnovitsky et al., 2010), recombinant Caf1A plug or soluble middle domain (PDB:3FCG, 2.85 Å) (Yu et al., 2009) and recently subunit-free Caf1M (PDB:4AY0, 1.52 Å) as well as the Caf1A N-terminal domain alone (PDB:4B0E, 2.0 Å) and in complex with Caf1M:Caf1 (PDB: 4B0M, 1.8 Å) (Di Yu et al., 2012). The only remaining unsolved structure in the Caf system is the Caf1A usher barrel. As reviewed in the introduction, the most extensively studied fimbrial usher proteins are type I pili (FimD) and P pili usher (PapC). The structure of PapC is in the closed state (PDB: 2VQI,

3.2 Å) and for FimD, structures are available for the closed usher (PDB: 3OHN, 3.01 Å) and in two different open states; one is FimD:FimC:FimH (PDB: 3RFZ, 2.8 Å) accommodating FimH adhesion in the pore with solved resolution of 2.8Å and the other subunit complex is FimD:FimC:FimG:FimH (PDB: 4J3O, 3.8 Å) accommodating FimG in the pore, with resolution of 3.80Å.

4.4.1 Modelling of closed and opened Caf1A

To start modelling the Caf translocon, a 3D model of Caf1A usher was produced using IntFOLD2 and I-TASSER servers. IntFOLD2 was initially used to model the open Caf1A based on FimD with bound FimC:FimH (chaperone:adhesin) complex and the FimH adhesin located within the pore (3RFZ:B) Figure 4.4. As IntFOLD2 cannot model a protein based on the assigned template, the alternative way is to use I-TASSER. Therefore, I-TASSER was used to build closed Caf1A based on 3OHN:A and opened Caf1A based on 4J3O:D (FimD:FimC:FimG: FimH complex with FimG located in the pore).

The C-score was used to estimate the quality and confidence level of each model. The C-score in I-TASSER ranges from -5 to 2 where the higher score signifies a model with a high confidence. The C-score for I-TASSER modelled closed Caf1A was 0.50 and for the open Caf1A model based on FimD accommodating FimG, the score was -0.23. In IntFOLD2, model quality is assessed by global model quality score where the score ranges between 1 and 0. A model with a score of less than 0.2 indicates the model might be incorrect and a score more than 0.4 indicates that the model is more complete and confident. The global model quality score of the Caf1A based on FimD accommodating FimH was 0.6928. These results signified that while all Caf1A models are confident, the open Caf1A based on FimD

accommodating FimH was more reliable than that with FimG. This may be related to the lower resolution (3.80Å) of the crystal structure of FimD with FimG which was used as template.

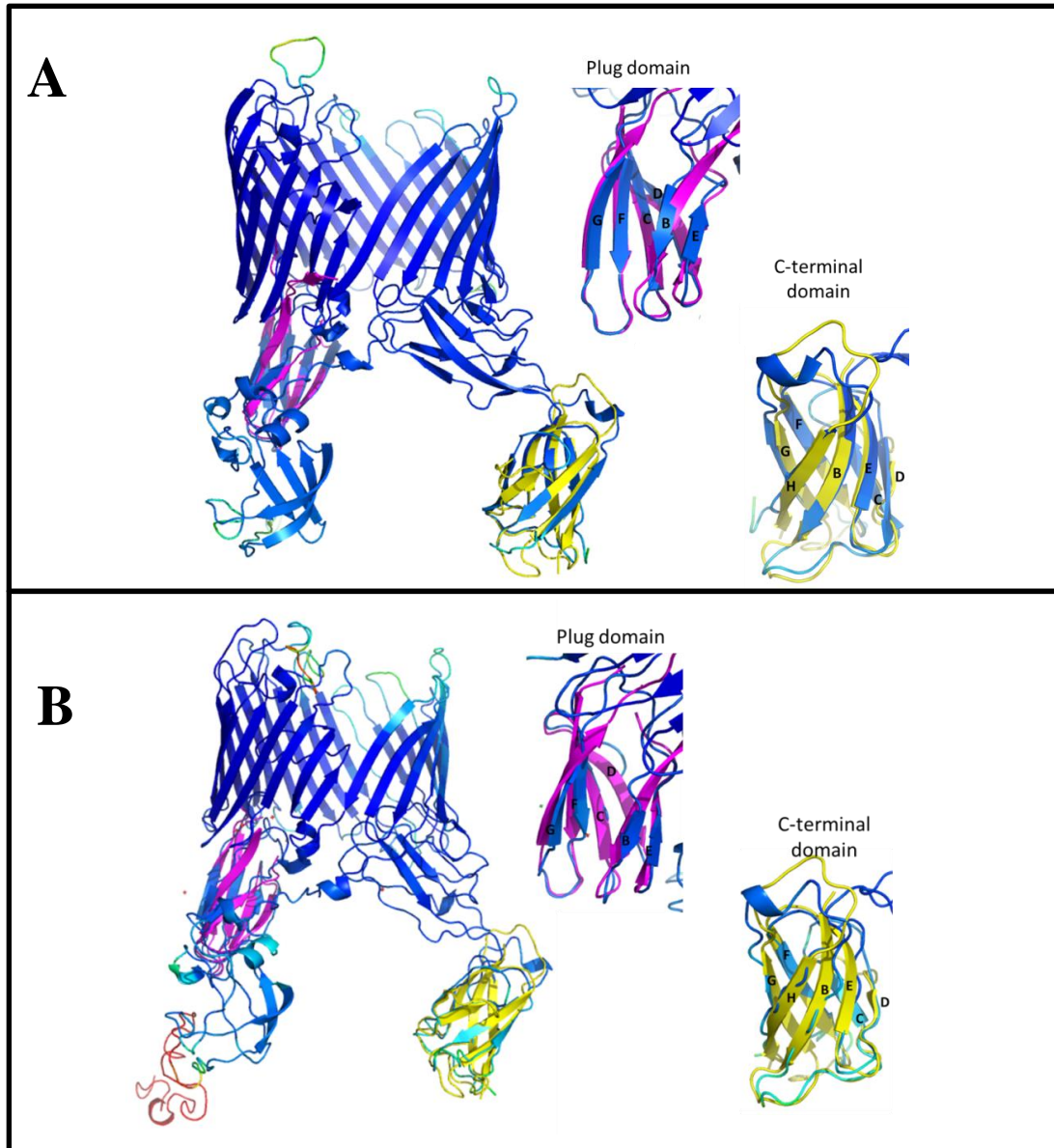


Figure 4.4: Quality and reliability assessment of Caf1A models based on ModFOLD4. The colour of the models were based on B-factor where blue indicates residues closest to the native structure and red residues furthest from the native structure (temperature scheme). (A) Caf1A model based on FimD accommodating FimH (IntFOLD2). (B) Caf1A model based on FimD accommodating FimG (I-TASSER). Yellow indicates solved C-terminal domain structure (PDB: 2XET, 1.6 Å) and pink indicates solved plug domain structure (PDB: 3FCG, 2.85 Å) superposed on the model structure.

To further assess the quality and reliability of the models, ModFOLD4 was used. With ModFOLD4, quality of the model is reflected in colour of the modelled barrel. The B factor of Caf1A based on FimD accommodating FimH (total barrel β strands are 24, in both) is better than that of Caf1A based on FimD accommodating FimG (total barrel β strands of Caf1A are 23, where one β -strand at β 19-strand is not assigned - see Figure 4.5 (C) below). The whole Caf1A model based on FimD accommodating FimH is in blue, indicating that the model is close to the native structure, but the C-terminus of the Caf1A model based on FimD accommodating FimG was coloured in cyan. Figure 4.4 shows model to structure superpositions for Caf1A barrel models and the Caf1A plug domain crystal structure. Caf1A model based on FimD accommodating FimH superposition has an RMSD=1.070 where the β -strands were aligned properly in the model, indicating that the model is a close representation of the native structure (see Figure 4.4 (A)). For Caf1A model based on FimD accommodating FimH superposition has RMSD=0.912 and all β -strands of the structure and the model were aligned.

The Caf1A model based on FimD accommodating FimG superposition has an RMSD=1.128 (PyMOL align) where one β -strand was not assigned in the model (strand-D) and β 1-strand is very short (see Figure 4.4 (B)). For the C-terminal domain of the Caf1A model based on FimD accommodating FimG (see Figure 4.4 (B)), superposition with the structure of Caf1A-C has RMSD=1.174 with four unassigned β -strands (strand-B, -C, -E and -H). Thus, because the Caf1A model based on FimD accommodating FimH appears to be more reliable than Caf1A model based on FimD accommodating FimG, it was used to assemble the translocon.

As reviewed in the introduction, for FimD the dimensions of the pore vary substantially between opened and closed pore (Phan et al., 2011). To predict the size of the pore of predicted Caf1A protein at different conditions, the Caf1A models were superimposed. As would be expected, this was also the case for modelled Caf1A (Figure 4.5). Dimensions changed from 27.4 Å by 55.2 Å in the closed modelled pore of Caf1A to 52.7 Å by 34.9 Å in the opened pore based on FimD accommodating FimH and 51.2 Å by 37.2 Å in the opened pore based on FimD accommodating FimG. As shown in Figure 4.5 (D), there is some conformational change on the predicted model of Caf1A based on FimD accommodating FimH (coloured in cyan) and FimG (coloured in pink) involving strand $\beta 1$, $\beta 2$, $\beta 3$, $\beta 4$ and $\beta 24$. This might be because the Caf1A model based on FimG is not reliable. However, the conformational changes between closed and open pore suggest flexibility within the barrel. Flexibility of the barrel is a promising feature for utilisation of the F1 translocon as a scaffold for export of heterologous peptides.

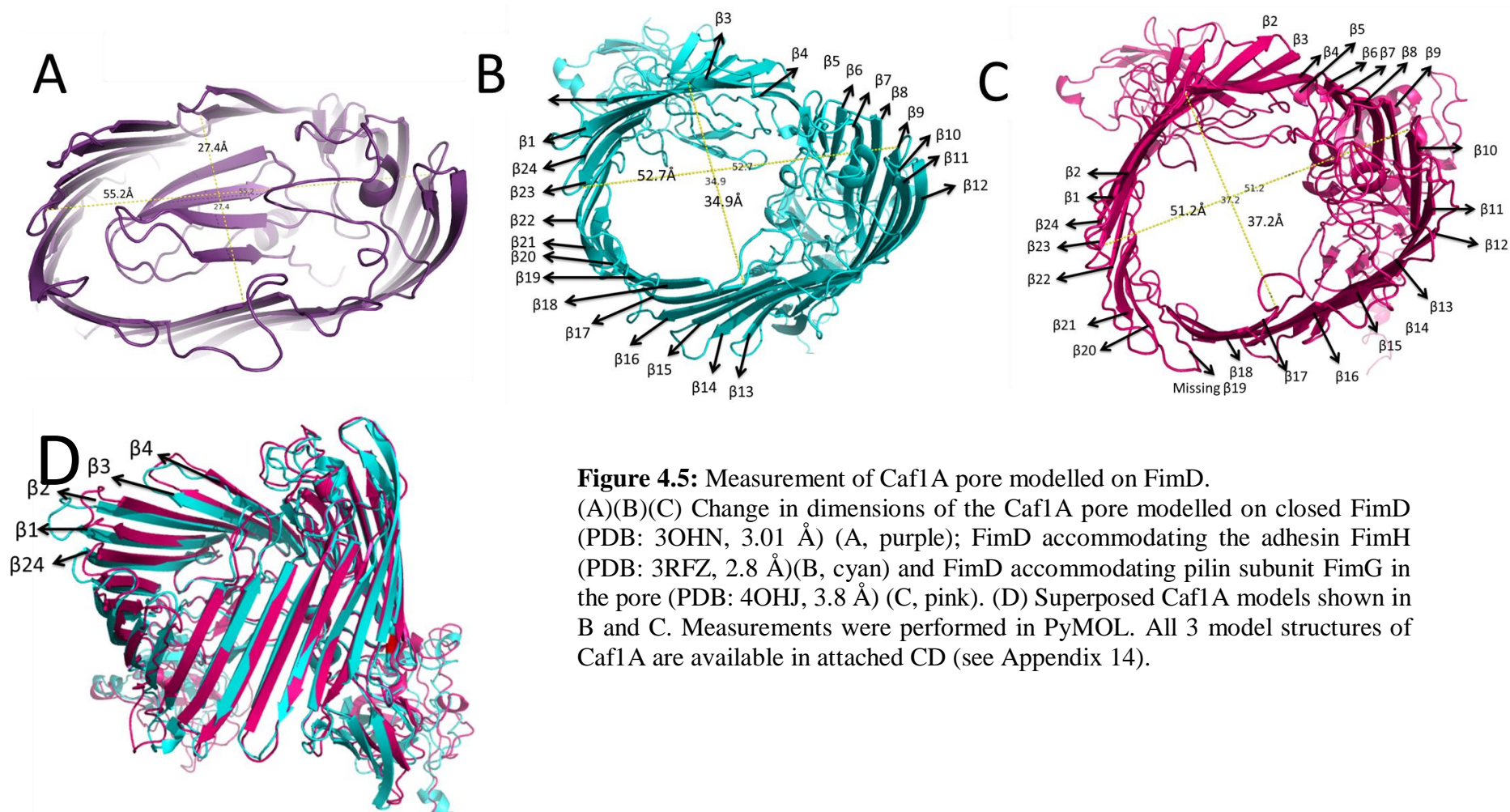


Figure 4.5: Measurement of Caf1A pore modelled on FimD.

(A)(B)(C) Change in dimensions of the Caf1A pore modelled on closed FimD (PDB: 3OHN, 3.01 Å) (A, purple); FimD accommodating the adhesin FimH (PDB: 3RFZ, 2.8 Å)(B, cyan) and FimD accommodating pilin subunit FimG in the pore (PDB: 4OHJ, 3.8 Å) (C, pink). (D) Superposed Caf1A models shown in B and C. Measurements were performed in PyMOL. All 3 model structures of Caf1A are available in attached CD (see Appendix 14).

Having modelled Caf1A, Caf1M:Caf1:Caf1 complex was assembled into Caf1A using TMalign and PyMOL. To start, TMalign was used to carry out superposition of the Caf1A based on FimD, 3RFZ chain B (TM-score=0.91412, where TM-score between 0.5 and 1 indicates same fold of structural similarity and less than 0.3 indicates random structural similarity (Zhang and Skolnick, 2005)). This step is crucial to obtain model coordinates relative to the template. Then, Caf1M (1Z9S chain A) was superposed with FimC (3RFZ chain C) to gain coordinates for Caf1M relative to FimC (TM-score=1.0). Then, the models of Caf1A, Caf1M:Caf1:Caf1 complex structure (PDB:1Z9S, 2.2 Å) and TMalign files were loaded into PyMol to view the models for assembly. In PyMol, Caf1A model was aligned with the TMalign file of Caf1A based on FimD. The Caf1M:Caf1:Caf1 complex (PDB: 1Z9S, 2.2 Å) was then aligned with the TMalign file of Caf1M based on FimC. The completed translocon (Figure 4.6) was saved as *.pse format for viewing and manipulation for later use (available in attached CD). Construction of the translocon using Caf1A model based on FimD accommodating FimG was also attempted. However, there was a clash between Caf1 and this Caf1A model even on trying to dock the subunit and chaperone one by one. This might be a consequence of the poor model with a number of unassigned β -strands possibly due to the low resolution of the template, and inability of Pymol to assign these strands or both. Therefore, the F1 translocon modelled on FimD:FimC:FimH was used in all further studies.

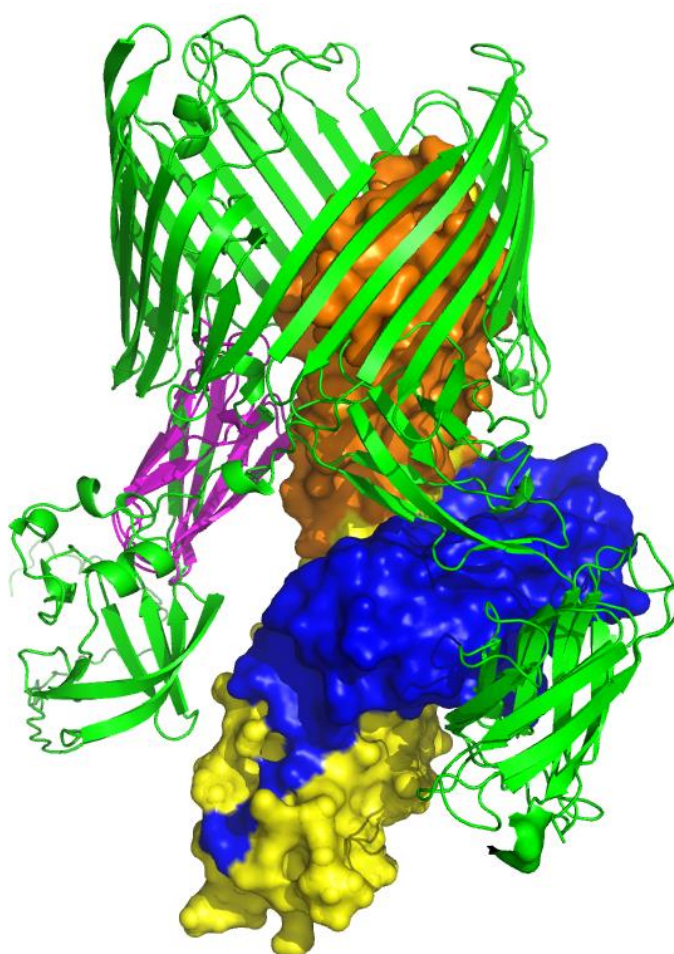


Figure 4.6: Model of assembled Caf translocon.

In green is Caf1A model, orange is top Caf1 subunit, yellow is bottom Caf1 subunit and blue is Caf1M chaperone. Caf1M:Caf1:Caf1 is solved crystal structure (PDB: 1Z9S, 2.2 Å). Pink, plug (PDB: 3FCG, 2.85 Å) in open position.

4.4.2 Modelling of *E. coli* SE11 translocon

The usher encoded in *E. coli* SE11 pSE11-2 was the only usher closely related to Caf1A. It shares 70% sequence identity with Caf1A. Therefore, this usher was also modelled to assist characterisation of Caf1A barrel. IntFOLD2 was used to model the usher based on multiple templates; 3RFZ:B (FimD), 2XET:A (Caf1A CTD), 3BWU:D (FimD NTD). The global model quality score of the usher based on FimD accommodating FimH was 0.6446, a value just marginally lower than that for the Caf1A model, which signifies a

high confidence model. In the model of this *E. coli* SE11 usher, the barrel was composed of 24 β -strands with a similar topology to the FimD structure. The superpose of *E. coli* SE11 usher, Caf1A and FimD revealed slight differences in the modelled conformation of β -strand 1, 2, 4, 21, 22, 23, and 24 (See Figure 4.7). Interestingly, these are the β -strands close to the subunit during translocation of polymer.

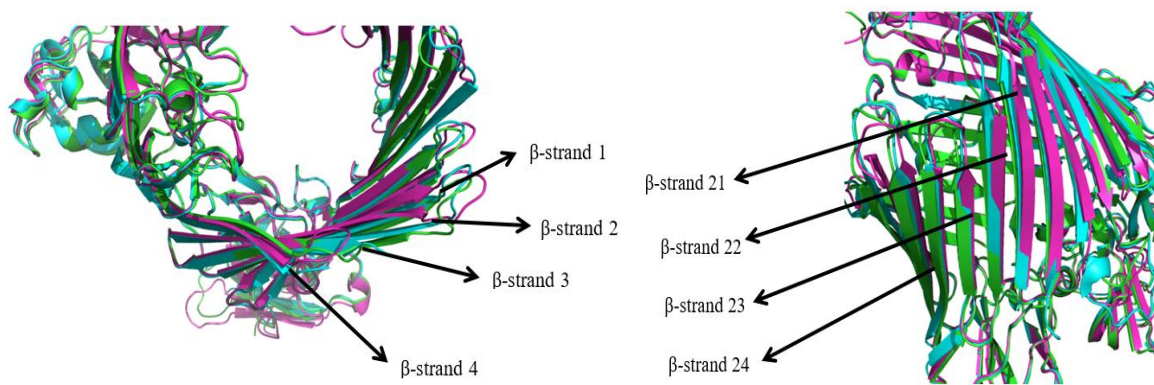


Figure 4.7: Superposition of modelled *E. coli* SE11 usher (pink) and Caf1A (green) with structure of FimD (blue).

4.5 Characterisation of Caf barrel

4.5.1 Residues interacting with conserved residues in the plug domain

The usher models of *Y. pestis* Caf1A and *E. coli* SE11 usher were used to localise residues that might be directly involved in F1 assembly and in particular the nature of residues facing the translocating subunit in the barrel. Previous studies had identified residues in the plug that are conserved within all fimbriae usher families. Selected conserved residues between Caf1A and PapC middle domain were shown experimentally to have an impact on F1 assembly when mutated (Mycroft, 2011; Yu et al.,2009). The translocon model permits analysis of how both plug and subunit might interact with the usher barrel. Multiple sequence alignment (MSA) of the mature usher of Caf1A, FimD,

usher from *E. coli* SE11, and other members of the $\gamma 3$ fimbrial usher protein family identified conserved residues (Figure 4.8). There are a total of 14 conserved residues within the plug domain (Figure 4.8).

The translocon models were then used to identify conserved residues in the plug that interact with either Caf1 subunit or with the β -barrel in the closed usher. An absolutely conserved Tyr residue (Tyr266) that lies at the beginning of β -strand D of the plug, was found to make a polar interaction with subunit Asn81 with a distance of 3.0 Å. (See Figure 4.8 (B)). Asn81 is notable for its conservation between Caf1 and SE11 subunits (see Chapter 6).

Glu298 was also absolutely conserved in the alignment and mutation of Glu298 had previously been shown to have a negative impact in production of Caf1 (Mycroft, 2011). In this previous study, mutation of E298K and E298R, led to a 2.9 and 2.3 times increase, respectively, in accumulated periplasmic Caf1. Arg263 was found to interact with the barrel. Arg263 is located at the end of β -strand C in a highly-conserved R/KQ sequence, but which is different in both Caf1A (LR) and the SE11 usher (QQ). In Caf1A Arg263 interacts with Tyr215 in the hairpin loop at a distance of 3.2 Å. Based on the crystal structure of closed PapC, Remault and co-workers (Remault, 2008), suggested that residues residing in the hairpin loop interact with and hold the plug in place. Previous mutation of R263D increased accumulation of periplasmic Caf1, 1.77 times and R263E increased accumulation of 2 times and 2.3 times in periplasmic and surface respectively (Mycroft, 2011). This suggested that these residues may not only be essential in correct positioning of the plug in closed usher, but also in correct orientation in the open usher.

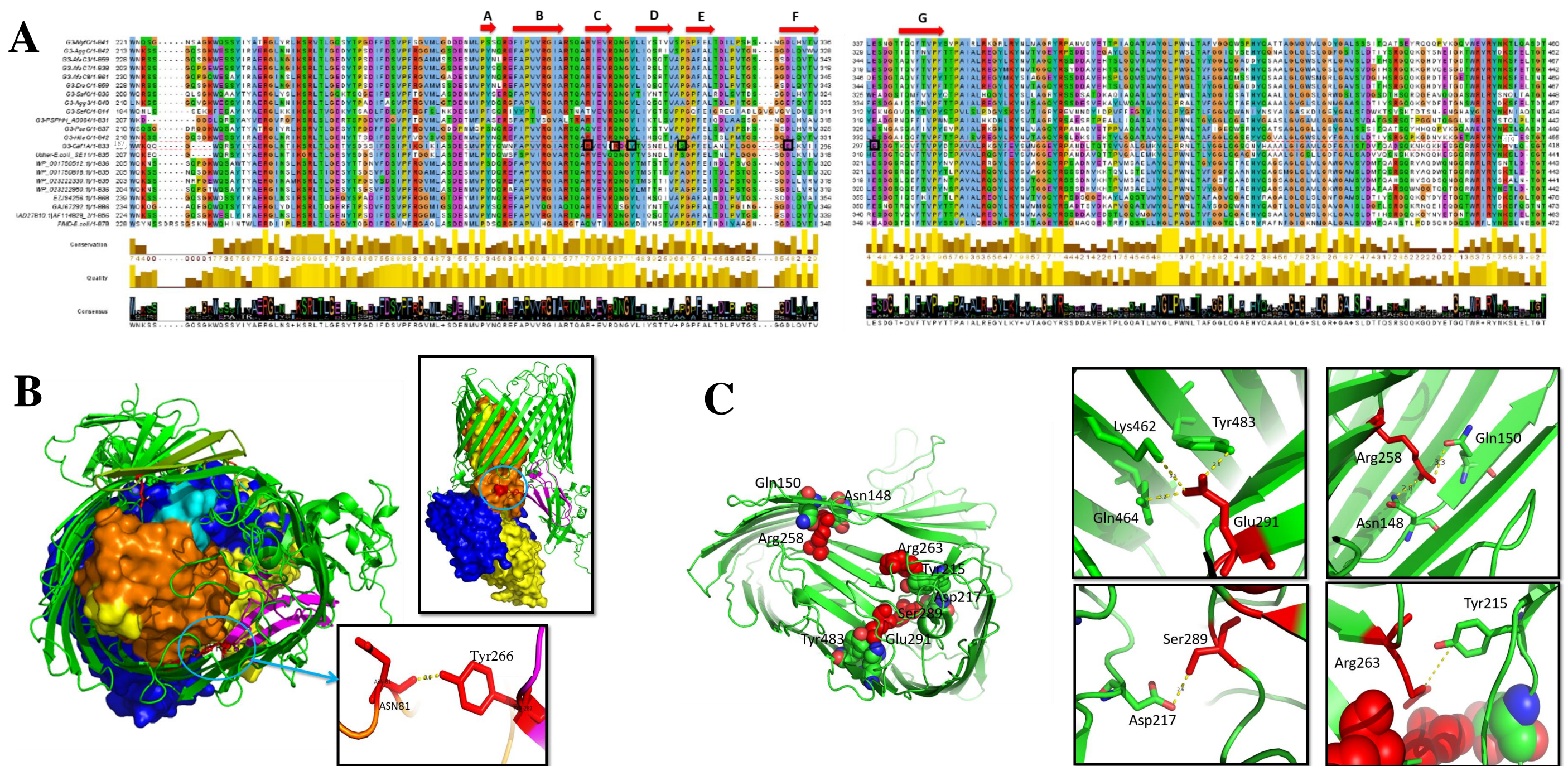


Figure 48: Multiple sequence alignment and interaction of middle domain (plug) with β -barrel and subunit.
 (A) Multiple sequence alignment of $\gamma 3$ fibrillar usher protein with focus on plug region, plug β -strands indicated with red arrow. Black boxes indicate highly conserved region in between strand C to strand F of plug domain modelled to interact with Caf1A barrel. For whole alignment see Appendix 2. FimD was also added in this alignment for comparisons. Identified conserved region is boxed in black. Jalview was used to visualise the alignment. (B) Interaction between subunit Asn81 with plug domain Tyr266 in opened Caf1A. (C) Interactions of plug residues with barrel residues, respectively, in closed Caf1A; Glu291 with Tyr483, Lys462 and Gln464; Ser289 with Asp217; Arg263 with Tyr215; and Arg258 with Asn148 and Gln150 in closed Caf1A.

Some additional non-conserved residues were also found to interact with β -barrel residues in the closed Caf1A. As shown in Figure 4.8 (C), Glu291 (a conserved negatively charged residue within β -strand E) was found to interact with barrel residues Tyr483, Gln464 and Lys462 with a distance of 2.8Å, 3.2Å and 3.5Å respectively. Arg258 (beginning of β -strand C) is mainly conserved within γ 3 family ushers, but not in FimD (γ 1). This residue interacts with Asn148 (barrel) and Gln150 (barrel) with a distance of 3.3Å and 2.8Å, respectively, while the non-conserved Ser289 (between β -strand D and -E) interacts with Asp217 in the hairpin loop with a distance of 2.6Å. As all six residues identified above might be important in correct localisation of the plug, these were included in the mutagenesis strategy (below) to determine if there was any impact on production and assembly of Caf1. The two residues, Arg263 and Glu298 that were identified as important in the previous study (Mycroft, 2011) were also mutated in this study for comparison of the impact of all mutants together using the pACYCF1 vector.

4.5.2 Localisation of key residues within Caf1A barrel

Analysis of the barrel started with prediction of the topological properties of the usher cavity. A programme called BetaCavityWeb was used to predict the cavity using the modelled *E. coli* SE11 usher and Caf1A structures and the solved crystal FimD structure. The programme cannot calculate cavity volume; other information was used to interpret usher cavity. Based on the results (depicted in Figure 4.9), FimD has 365 atoms that contribute to the channel boundary, Caf1A has 370 atoms and *E. coli* SE11 usher has 330 atoms. The most important information that this programme produced was the ‘neckbottle’ where it gives a value of the cavity at the narrowest point. At the narrowest cavity, Caf1A

has a pore width of only 32Å, while the narrowest measurement is 34Å for the *E. coli* SE11 usher and 40Å for FimD. This suggests that Fim has a larger cavity due to the size of the terminal adhesin, FimH which contains two domains and may require more space for export compared to Caf subunit. Although there is no structure of the *E. coli* SE11 subunits yet, these preliminary results suggest that *E. coli* SE11 subunits may be of similar dimensions to the Caf1 subunits and that γ 3 family ushers in general may have a slightly smaller channel.

The cavity of the closed Fim barrel, that would accommodate the plug domain, was also checked to compare the size of the narrowest cavity with that of the open usher. The result shows that closed FimD has 321 atoms that contribute to the channel boundary which is lower than in the open FimD structure. The size of the narrowest cavity was only 30Å consistent with the proposal (Geibel et al., 2013) that ushers have a flexible barrel structure. An usher with flexible cavity dimensions that do not change the fold or export function of the usher, is ideal for applications of Caf1A as a scaffold for export of short peptides.

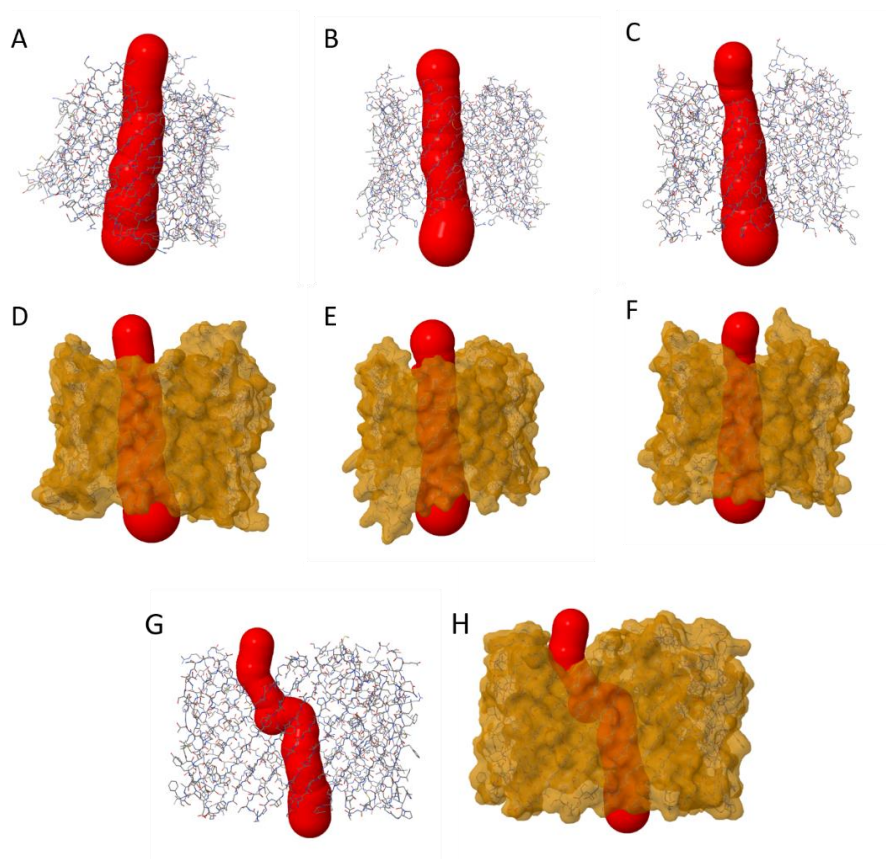


Figure 4.9: Usher cavity predicted by BetaCavityWeb.

Cavity for (A)(D) model of open Caf1A barrel, (B)(E) open FimD barrel structure from 3RFZ (2.8 Å) (C)(F) model of open *E. coli* SE11 usher model and (G)(H) closed FimD structure. Red indicates predicted cavity in the barrel and (A)(B)(G)(H) show β -barrel in stick and (D)(E)(F)(H) show β -barrel in surface.

In order to identify residues facing the barrel cavity that might be important for translocation and polymerisation of Caf1, the barrel segments of Caf1A, *E. coli* SE11 usher and FimD were analysed by MSA for conserved residues which were then located within the modelled Caf1A usher (Fig 4.10). This was done to identify conserved residues that might be required for interaction with fibre subunits within the barrel. Figure 4.10 (B) shows all residues facing inwards throughout the β -barrel and (Figure 4.10) (A) and (C) highlight the 33 conserved residues facing inwards with red highlight and red dots, respectively. The Caf translocon model was also used to identify any modelled interactions of these residues

with Caf subunit. In total eight of these 33 conserved residues (six in the β -barrel wall and 2 in the hairpin loop) were modelled interacting with Caf1 or the plug. Six of these residues Tyr215, Asp217, Asn148, Gln150, Lys462 and Gln464 interact with the plug in the closed position as shown in Figure 4.8. Interestingly, residues Asn148 and Gln150 (both located in β 1) and shown to interact with plug residue Arg258 (Figure 4.8 C), are also modelled interacting with 2 subunit residues in the open usher during Caf1 assembly. Caf1 subunit Thr166 interacts with Caf1A Asn148 and Caf1 Thr53 interacts with Caf1A Gln150 at a distance of 2.7 Å and 5.2 Å respectively (see Figure 4.10(C)). The two additional conserved residues of Caf1A barrel, Asn144 and Asn601 also interact with subunit in the modelled translocon. Asn144 is located in Caf1A β 1 and interacts with Caf1 loop residue Ser93 with a distance of 1.7 Å. Asn601, located in Caf1A β 22, interacts with subunit, Ser168 with a distance of 2.9 Å. As these residues (summarised in Table 4.3) might have an important role in F1 assembly each residue was independently mutated and the impact on production of Caf1 assessed.

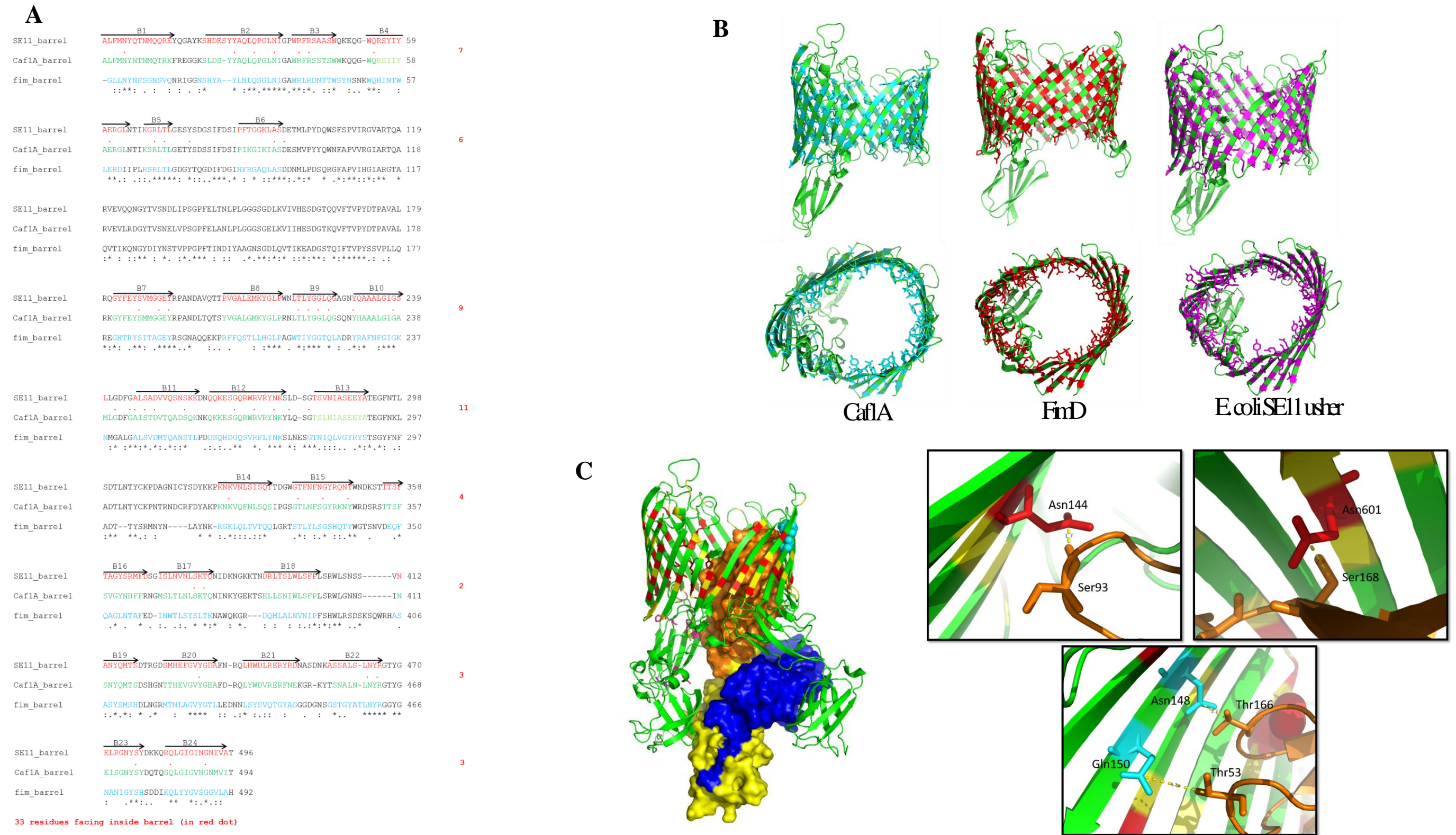


Figure 4.10: Analysis of important residues in Caf1A β -barrel.

(A) Multiple sequence alignment of barrel and plug domains of *E. coli* SE11 usher, Caf1A and FimD. Black arrows indicate β -strands forming the barrel. Numbering starts at the first residue of β 1 strand of the usher barrel; therefore, add 139 to account for residues of the N-terminal domain and adjust residue number to that of intact mature Caf1A as used in images. Asterisk (*) indicates conserved residues and red dot indicates conserved residues facing inwards towards the barrel cavity. (B) Modelled Caf1A, *E. coli* SE11 usher and FimD structure highlighting alternating residues facing inside the barrel, labelled turquoise, red or purple. (C) Caf translocon, showing conserved residues, labelled in red (residues facing inwards) and yellow (residues facing outwards). Cyan spheres are Caf1A Asn148 and Gln150 that are modelled interacting with Caf1 subunit during F1 assembly and also with the plug in closed Caf1A usher (see Fig 4.8 C). In green is Caf1A model, orange is top subunit, yellow is bottom subunit and blue is Caf1M. The numbering for the alignment was exclude the N-terminus sequences. Therefore, add 139 to get the correct number of residue as in the model.

Residue	Location	Interaction in opened model	Interaction in close model
Glu291	Plug	-	Caf1A_Lys462 (β 14), Caf1A_Gln464 (β 14) and Caf1A_Tyr483(β 15)
Glu298	Plug	-	-
Arg258	Plug	-	Caf1A_Asn148 (β 1) and Caf1A_150 (β 1)
Tyr266	Plug	Caf1_Asn81	Caf1A_Tyr198 (β 4) and Caf1A_Tyr215 (β -hairpin)
Arg263	Plug	-	Caf1A_Tyr215 (β -hairpin)
Ser289	Plug	-	Caf1A_Tyr217 (β -hairpin)
Tyr215	Hairpin	-	Caf1A_Arg263 (plug)
Asp217	Hairpin	-	Caf1A_Ser289 (plug)
Asn144	Barrel (β 1)	Caf1_Ser93	-
Asn148	Barrel (β 1)	Caf1_Thr166	Caf1A_Arg258 (plug)
Gln150	Barrel (β 1)	Caf1_Thr53	Caf1A_Arg258 (plug)
Lys462	Barrel (β 14)	-	Caf1A_Glu291 (plug)
Gln464	Barrel (β 14)	-	Caf1A_Glu291 (plug)
Asn601	Barrel (β 22)	Caf1_Ser168	-

Table 4.3: Summary of Caf1A residues with potential importance on Caf translocon function.

4.6 Characterisation of important residues in Caf1A by mutagenesis

As discussed above, six residues within the plug were modelled as being involved in interactions that might be important to usher function and export of F1 polymer (see Table 4.3 and section 4.5.1). These residues, Arg258, Arg263, Tyr266, Ser289, Glu291 and Glu298, were mutated individually to the smaller hydrophobic residue, Ala. This change would disrupt any specific interaction formed by these charged and polar hydrophilic residues. Site-directed mutagenesis (Quickchange II XL kit) with pACYCF1 as a template, was used to construct the mutated usher within the complete *caf* locus. As described, (see section 2.4.4) mutated plasmid DNA was initially recovered and sequenced from *E. coli* Top10 and confirmed mutated plasmid then

transformed into *E. coli* BW25113 for analysis and quantitation. F1 production was induced at 37°C for 4 hours, surface F1 recovered by heat shock at 65°C for 1 hour and the level of F1 polymer produced quantitated following heat denaturation and SDS-PAGE.

4.6.1 Investigating the role of Tyr266 and Glu298

One of the residue among the highly conserved residues in multiple sequence alignment of $\gamma 3$ fimbrial usher protein is Tyr266. This plug residue is not only conserved but was also modelled to interact with Asn81 in the subunit when the usher is in the active translocating state (see Figure 4.8 (B)). Notably, this Asn81 is also adjacent to a conserved GlyAsn in alignment of Caf1 and the *E. coli* SE11 CU subunits (See Chapter 6 below). This suggests that the plug residue Tyr266 might play a role in stabilising F1 translocation via transient interaction with Caf1. Two point mutants within Caf1 subunit were created. In Y266S, Tyr was replaced by a smaller polar residue and in Y266A by a small hydrophobic residue that would interfere with any natural interaction that Tyr266 might form. As showed in Figure 4.11 Tyr266 is not absolutely required for F1 assembly. There might be an indication of decrease in surface assembly of F1 with Y266A, but confirmation of this would requires analysis and quantitation of triplicate samples.

The closed Caf1A model was also assessed to look for any interaction of Tyr266. When PapC is superposed with Caf1A, Caf1A Tyr266 corresponds to PapC Arg281. This PapC plug residue (Arg281) is predicted to undergo repulsion with Arg210 in the β -barrel (Remaut et al., 2008). When checked in the Caf1A closed model, Tyr266

interacts with Tyr195 (corresponding to Arg210 in PapC) and with Tyr215 in the β -hairpin (Figure 4.12). These extensive hydrophobic interactions between plug, β -barrel and β -hairpin form a hydrophobic patch and might be important in positioning the plug inside the pore. Mutation of Tyr 215 might be predicted to destabilise the triple Tyr interaction. With this mutation the F1 level was slightly reduced to $86.37\% \pm 9.09$ compared to wildtype (see Figure 4.14). Ideally, impact of these residues would be further investigated by double and triple mutation of all three Tyr to test the impact on F1 assembly and Caf1A assembly and permeability.

Glu298 is also conserved throughout the $\gamma 3$ fimbrial usher. Glu298 was mutated to Lys (large and oppositely charged residue) in the previous study, using the pFMA1 expression system (Mycroft, 2011). This mutation resulted in an accumulation of 2.87-fold Caf1 in the periplasmic compared to 5.4 fold of the N-terminal deletion mutant which acted as a non-exporting control. It was concluded that Glu298 is important in translocation of F1 but not critical. As pACYCF1 is a much more efficient system, Glu298 was mutated to Lys to test the impact in pACYCF1. As shown in Figure 4.14 below, the level of F1 in E298K was decreased to $84.03\% \pm 6.42$, consistent with the observed accumulation of Caf1 in the pFMA1 system carrying the E298K mutation. This new system, pACYCF1 confirmed an effect on translocation of F1 to the cell surface while the earlier pFMA1 system, recorded Caf1 accumulation in the periplasm due to overexpression of chaperone.

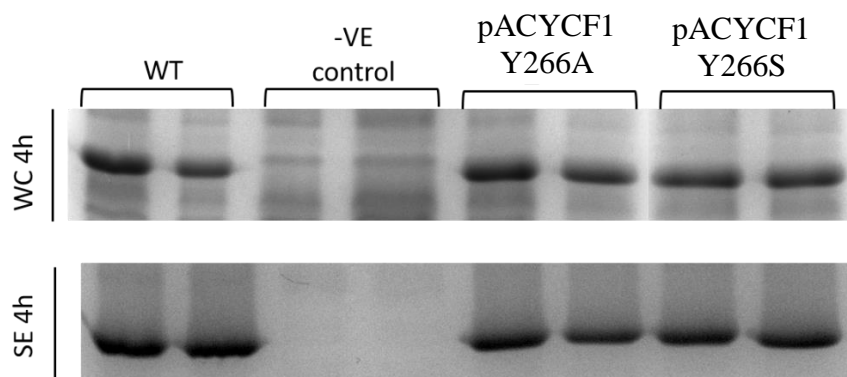


Figure 4.11: SDS-PAGE of surface F1 extracted from *E. coli* BW25113 from pACYCF1_Y266A and pACYCF1_Y266S constructs. 4 OD cells (for whole cells) and 3 OD cells (for surface extraction) after 4 hour thermoinduction. For surface extraction cells were suspended in 100 μ l PBS. 10 μ l of samples were loaded onto each lane. Two separate transformants processed for each construct.

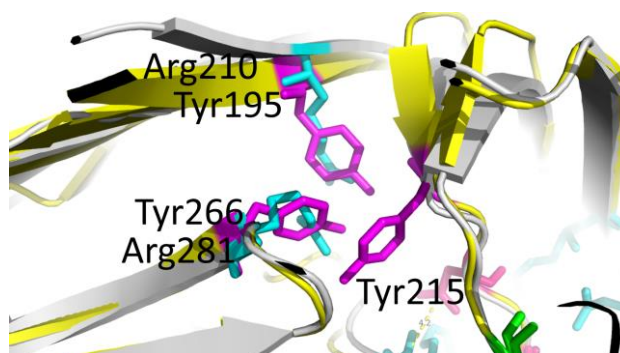


Figure 4.12: Superpose of closed PapC structure (3FIP, white) and Caf1A closed model (yellow). Interaction between Arg281 (plug) and Arg210 (β -barrel) in PapC - residues coloured in cyan. Interaction between Caf1A Tyr266 (plug), Tyr195 (β -barrel) and Tyr215 (β 5-6 hairpin) - residues coloured in pink.

4.6.2 Mutation of residues predicted to be involved in pore gating

Four additional plug residues (Arg258, Arg263, Ser289, and Glu291) were found to interact with the β -barrel in the closed Caf1A model suggesting that these residues may contribute to holding the plug in the pore (refer figure 4.8). These four residues were mutated to Ala to disrupt any interactions that occurred naturally in the

translocon. R263A was shown to have a slight increase the level of F1 ($111.71\% \pm 6.39$) compared to wildtype (Figure 4.14). Based on the translocon model, Arg263 interacts with Tyr215 from β 5-6-hairpin (between β 5 and β 6) (refer Figure 4.10). Mutation of Arg263 to Ala, a hydrophobic residue that will bury inside the protein core, should disrupt the predicted interaction between Arg263 with Tyr215 in Caf system, thus potentially increasing opening of the pore. However, mutation of Tyr215 to Ala showed a decreased level of surface F1, $86.37\% \pm 9.09$ compared to wildtype Caf1A.

Mutation of R258A resulted in a similar level of surface F1 ($95.71\% \pm 7.71$) compared to wildtype (refer Figure 4.14). In the translocon model, Arg258 interacts with Asn148 and Gln150. To validate this interactions is not critical, Asn148 and Gln150 were mutated individually to Ala. Quantitation of surface F1 of N148A and Q150A showed F1 level of $98.26\% \pm 9.44$ and $82.05\% \pm 8.05$, respectively compared to wildtype. Asn144, Asn148 and Gln150, however, were also interacting with subunit. Thus, loss of Gln150 might have an impact on translocating the F1 fibre, or simply correct orientation of the plug This suggests that Arg258 might not be important alone but that loss of this Arg might have a greater impact if mutated together with Gln150 as Arg258 is mostly conserved throughout γ 3 fimbrial usher system.

The plug residue Glu291, was modelled to have an interaction with the barrel residue Lys462, and was therefore mutated to the small uncharged residue Ala and also to the positively charged Lys. The mutation of E291K would be predicted to cause repulsion with Lys462 in the β -barrel and weaken the plug interaction leading to an opened usher, while the mutation E291A would completely abolish any interaction with Lys462 (refer Figure 4.10 for interaction of Glu291). These changes resulted in a slight

increase in F1 level of $116.85\% \pm 6.2$ (E291A) and $111.46\% \pm 2.87$ (E291K) compared to wildtype respectively. These results are consistent with weakening of the plug interaction facilitating removal of the plug and opening of the usher. Mutation of K462A also showed a slight increase in surface F1 level of $109.54\% \pm 3.29$ which further supports the suggestion that destabilisation of the Glu291-Lys462 interaction facilitates F1 export. Glu291 corresponds to Arg305 in PapC (Figure 4.13). Point mutation of R305A was shown to increase pore opening by disrupting the salt bridge interaction between Arg305 with Glu467 (Volkan et al., 2013). These results with Caf1A are consistent with an ionic interaction between Glu291 and Lys462 helping to hold the plug in position in the pore.

Based on the closed translocon model, Ser289 in the plug interacts with Asp217 from the β 5-6 hairpin. Mutation of Ser289 had the most profound effect on surface F1 level, of any of the single point mutations tested Figure 4.14. The level of surface F1 decreased to $70.81\% \pm 3.0$ and $91.03\% \pm 1.31$ with the Caf1A mutations, S289A and D217A respectively. The superpose of the closed Caf1A model and PapC located Caf1A-Ser289 corresponding to PapC-R303, a key residue that forms a salt bridge with D234 and contributes to correct positioning of the plug inside the pore (Figure 4.13) (Volkan et al., 2013). This suggests that Ser289 and Asp217 may similarly be important in orientation of the plug.

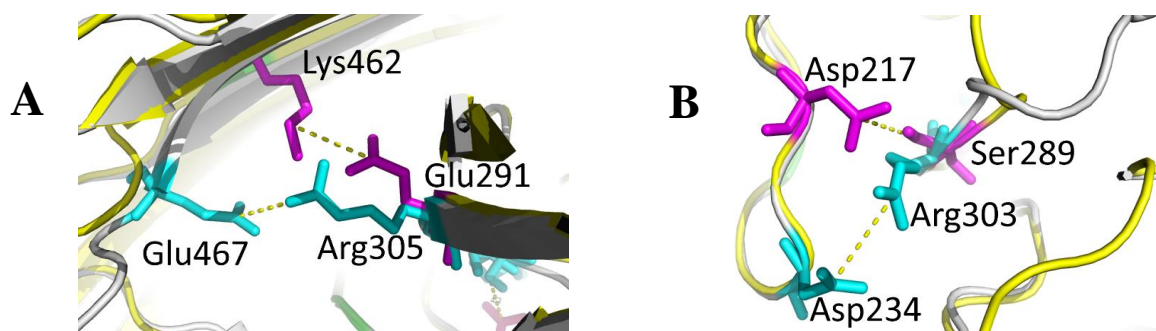


Figure 4.13: Superpose of closed PapC structure (3FIP, white) and Caf1A closed model (yellow).

(A) Interaction between Glu291 (plug) and Lys462 (β-barrel) in Caf1A. Glu291 corresponds to Arg305 in PapC. (B) Interaction between Ser289 (plug) and Asp217 (β5-6 hairpin) in Caf1A. Ser289 corresponds to Arg303 in PapC. Pink indicates residue from Caf1A and cyan from PapC.

4.6.3 Investigating the role of conserved residues in β-barrel

Based on the *in silico* analysis in section 4.5.2, two conserved residues in the barrel, Asn144 (located in β1) and Asn601 (located in β22) were identified as interacting with subunit. Single point mutations of N144A and N601A had no apparent effect on the level of accumulated surface F1 ($100.81\% \pm 5.97$ and $93.22\% \pm 7.37$ of wild-type levels of surface assembled F1, respectively). This suggests that single mutations within the barrel are likely not to have a great impact on the level of surface F1. Multiple point mutations would be more likely to have an impact.

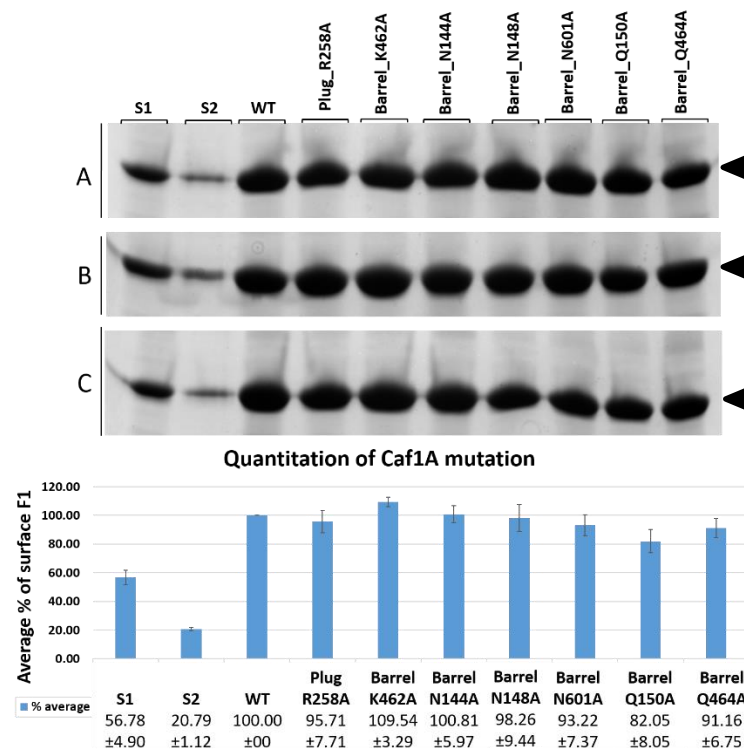
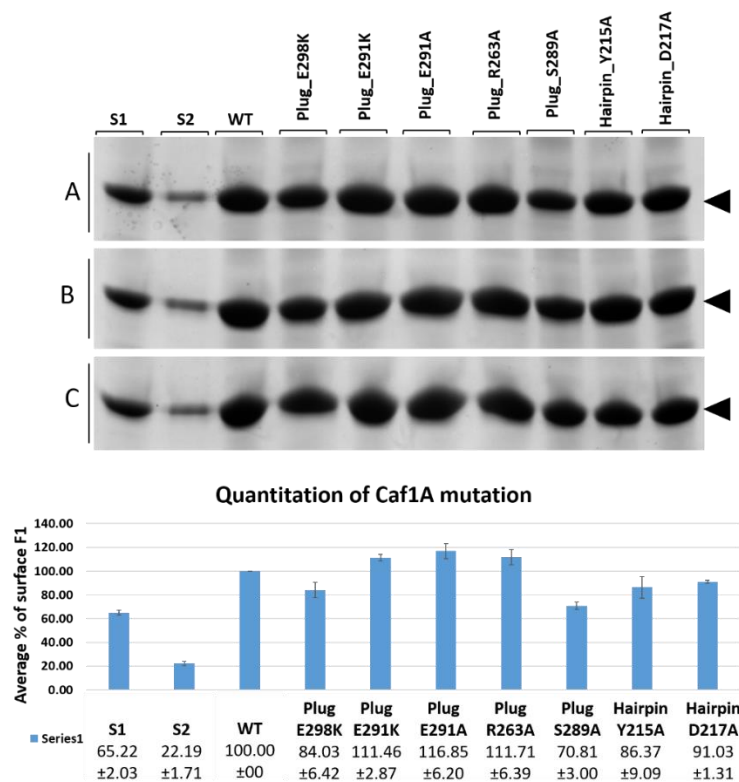


Figure 4.14: SDS-PAGE of extracted surface F1 from *E. coli*/pACYCF1 with point mutations in Caf1A.

Following 4 hour thermoinduction, F1 was extracted at 65°C for 1 h from 3 OD cells in 100µl PBS. Extracted samples (12µl) were mixed with 4µl SDS loading dye, heated at >95°C for 10 min to ensure complete denaturation of F1 fibre. Denatured sample (10µl) was loaded on SDS-PAGE. Following electrophoresis, protein was stained with Coomassie blue, imaged and bands of denatured Caf1 monomer (black arrowhead) quantitated. S1 and S2 indicate 2.59µg and 0.863µg respectively of the F1 standard applied. A, B and C are biological repeats, 3 separate transformants processed for each construct.

4.7 Conclusions

In this chapter, a plasmid encoded usher (pSE11-2) from *E. coli* SE11 was identified as a closely related usher to Caf1A with 70% sequence identity within 99% sequence coverage. A phylogenetic tree was constructed and this confirmed that the *E. coli* SE11 usher belongs to $\gamma 3$ group, and is the closest currently identified neighbour to Caf1A. This was validated with bootstrap value of 100% from 1000 replicates. The usher from *E. coli* SE11 and Caf1A are likely to share common functional features with respect to assembly via the chaperone-usher system, despite the fact that the *E. coli* usher assembles two different subunits while the Caf1A usher assembles only one. Despite the high similarity between *E. coli* SE11 usher and Caf1A, this usher could not be used as template for homology modelling of Caf1A as there is no solved structure. Therefore, PapC (PDB: PDB:2VQI, 3.2 Å) was used to model closed Caf1A and the structures of FimD:FimC:FimH accommodating FimH within the usher (PDB: 3RFZ, 2.8 Å) and FimD:FimC:FimF:FimG:FimH accommodating FimG in the usher (PDB: 4J3O, 3.8 Å) were used to model open Caf1A. However, modelling of Caf1A based on FimD:FimC:FimG:FimH gave a poor model where β -strands were poorly predicted with missing β -strands in the plug domain, C-terminal domain and the β -barrel. Thus, for open Caf1A, the model based on FimD:FimC:FimH combined with the solved structure of Caf1M:Caf1:CAf1 (PDB: 1Z9S, 2.2 Å) was used for assembly of the Caf translocon.

Both closed Caf1A usher and the modelled Caf translocon together with multiple sequence alignments were used to identify any interactions that might be important in

the functions of Caf1A to export F1 polymer. Residues predicted to have interactions within Caf1A were mutated to test the impact on surface F1 assembly. One important residue is the plug residue Tyr266 as it was predicted to interact with subunit Asn81 in the active state of Caf1A and to form an extensive hydrophobic interaction with Tyr195 (β -barrel) and Tyr215 (β 5-6 hairpin) in the closed usher. Mutation of Tyr266 and Tyr215 individually each appeared to lead to slight reduction in the level of surface F1. Whether this is a consequence of loss of interaction with the subunit or interference with plug opening remains to be seen. As identified in Chapter 5 subunit Asn80 appears to be more important for F1 assembly. The triple Tyr interaction may be important in positioning the plug inside the pore or even as a subunit recognition site. With PapC several ionic interactions have been proposed to hold the PapC plug within the translocation pore (Remaut et al., 2008).

Superpose of PapC and Caf1A revealed that plug residues in PapC involved in these ionic interactions (Arg281, Arg303, Arg305 and Arg316) correspond with Caf1A plug residues also identified as interacting residues. PapCArg281 corresponds to Caf1ATyr266, PapCArg303 to Caf1ASer289 and PapCArg305 to Caf1AGlu291. Among these residues, only Glu291 forms a typical salt bridge, forming an ionic interaction with the β -barrel residue Lys462 and loss of this interaction led to an apparent increase in the level of surface F1. This would be consistent with (i) closure of the plug within the barrel, via the ionic interaction between Glu291 and Lys462, being the primary role of this residue and (ii) that loss of this interaction facilitates opening of the plug without interfering with Caf1 assembly and in fact enhances translocation to the surface. Following polymerisation, as shown in Chapter 3, F1

remains anchored to the bacterial cell surface at least initially and this is assumed to be via its interaction with chaperone. Additional interactions of subunit with usher could stabilise surface association of F1. However, in this study extraction for quantitation was via heating at 65C for 1h in PBS. This completely denatures Caf1M releasing subunit and it is unlikely that any interaction with usher would have an impact on levels of F1 polymer released following this treatment. Thus, loss of anchoring interaction within the usher is unlikely to be the explanation for apparent enhanced recovery of surface F1 from cells expressing the E291A or K mutated usher. In contrast, mutation of Ser289 showed a significance reduction in surface F1. This could be due to a requirement for direct interaction with subunit or through an additional interaction within the open pore. Much more detailed investigation is required to establish involvement of these residues in opening and closing the pore. The models presented and preliminary mutagenesis provide a solid background for further study.

In this study, the residues that were tested were based on conserved residues facing inside the barrel and residues shown to have interaction in the translocon model. A systematic approach might involve mutation of all inward facing residues. Alternately, a mutagenesis based assay selecting for non-exporters would be another approach. However, none of the single mutations tested had a dramatic effect on the export of surface F1 and double mutations are likely to have a greater effect in highlighting key interactions. Therefore, a targeted approach focusing on interacting residues may remain a good direction for future study. This study also highlights flexibility of usher barrel residues facing the subunit, a fact used in the next stage of the investigation to optimise export of modified F1.

CHAPTER 5 CAF1 AS A SCAFFOLD FOR EXPORT OF HETEROLOGOUS EPITOPES

5.1 Introduction

E. coli and other Gram negative bacteria coat themselves in remarkably stable protein fibres assembled via the well-recognised chaperone-usher (CU) pathway. The simplest of these are long polymers, like the F1 fibres of *Yersinia pestis*. The F1 antigen, a polymer of the 15.6 kDa subunit Caf1 is produced at higher temperatures of 35-37°C and covers the bacterial surface making the bacteria more resistant to phagocytosis. Naturally, F1 fibre only disassociates into oligomers and polymers if exposed to a high temperature of 87.5°C (Zavialov et al., 2005). In 0.5% SDS, the fibre begins to dissociate at 75°C. This represents an appealing scaffold structure to display short foreign peptides. This chapter demonstrates the flexibility of the Caf CU pathway for surface display of different peptides and the ability to use the translocon model to identify problem areas and optimise export.

Earlier unpublished studies in this laboratory had investigated the flexibility of Caf1 subunit for periplasmic polymerisation following alteration of 4 different surface loops, using the pFM1 vector (expresses chaperone and subunit only, under ptrc control). In all cases, surface loops could be readily exchanged for Gly with little effect on polymerisation, but replacement with a short well-characterised epitope of HIV GP41 led to variable results (see Appendix 3 for detail). Transfer of the inserted sequences to the surface expression vector pFMA1 gave less clear results due to the lower level of surface F1 assembled with this system. Therefore, in this study, the ability to export foreign epitopes was tested using the pACYCF1 vector, under thermo-induction controlled by the native Caf1R regulator. Initial studies focused on requirement of native Caf1 sequences of these four surface loops for surface assembly.

Export of the short characterised viral epitope was then monitored. Having identified the optimum insertion site, the ability to decorate surface F1 with polyhistidine inserts was assessed. During this study, some key features of the usher and assembly process were also identified.

5.2 Prediction of B-cells epitopes in Caf1

To design vaccines and diagnostic kits, knowledge of B-cell epitopes are important because B-cell epitopes are the sites that are recognised by antibodies of the immune system. As B-cell epitopes are frequently located in loops they also represent short sequences that can be replaced without disruption of the core protein structure. For example, they might be replaced by heterologous epitopes to induce protection against other microbes. In 1988, Huges and Gillel induced protective immunity in mice against *P. aeruginosa* using linear B-cell epitopes (Gilleland et al., 1988).

In a previous study, four different loops of Caf1 that were not involved in β -strands had been replaced by either glycine residues or a charged epitope of HIV gp41 (Figure 5.1) to test the impact on subunit polymerization using the plasmid pFM1 (results are shown in Appendix 3). In this study, export of these modified Caf1 subunits through the translocon was tested using pACYCF1. DiscoTope Prediction (Kringelum et al., 2012) was used to review prediction of all linear B-cell epitopes of F1. Three loops (TIMDNGN (loop 1), TSQDGNNHQ (loop 4) and GSKGGKLAAGK (loop 7)) had strong epitope predictions, all of which had been included in the previous study (Figure 5.2 The second (KTG) and fifth loops (DSRDF) were also predicted as additional epitopes. However, the region between residues 100 and 120 were weaker

predictions and shorter sequences. Therefore, the four predicted epitope sites used in the previous study were also used here to assess surface assembly of F1 with mutated loop. Figure 5.3 shows locations of replacements with Gly or ELDKWA epitope within the Caf1 structure.

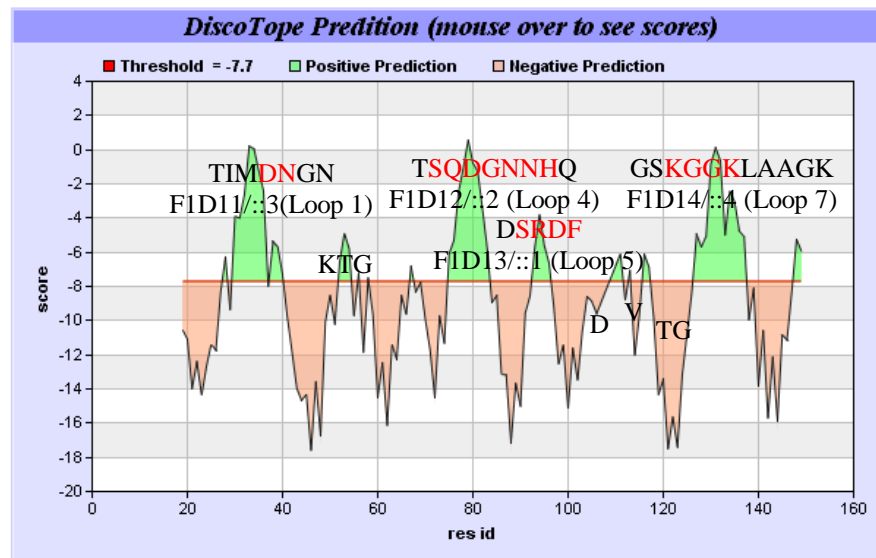
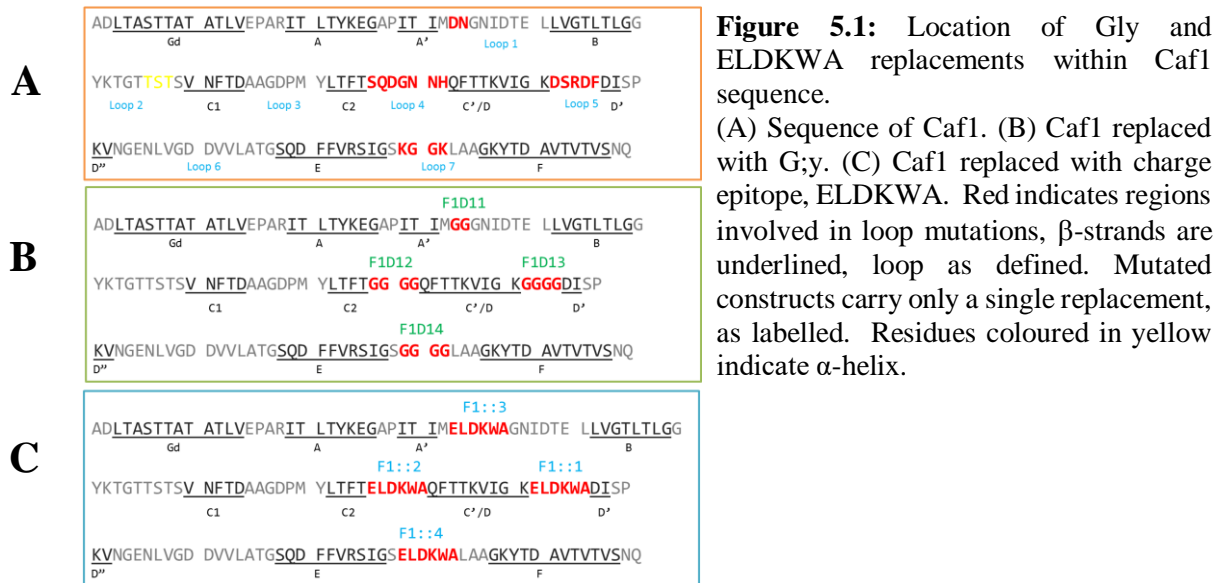


Figure 5.2: Epitope prediction in F1.

The output from DiscoTope Prediction programme, a structure based antibody prediction. In red are segments of each epitope previously identified and that were replaced by Gly residues or HIV gp41 core epitope, in pFM1 to monitor the impact of polymerisation. The default threshold, -7.7 was used where residues above the threshold were predicted as B-cell epitopes. 1Z9S (2.2 Å) (Caf1:Caf1:Caf1M) structure was used for prediction.

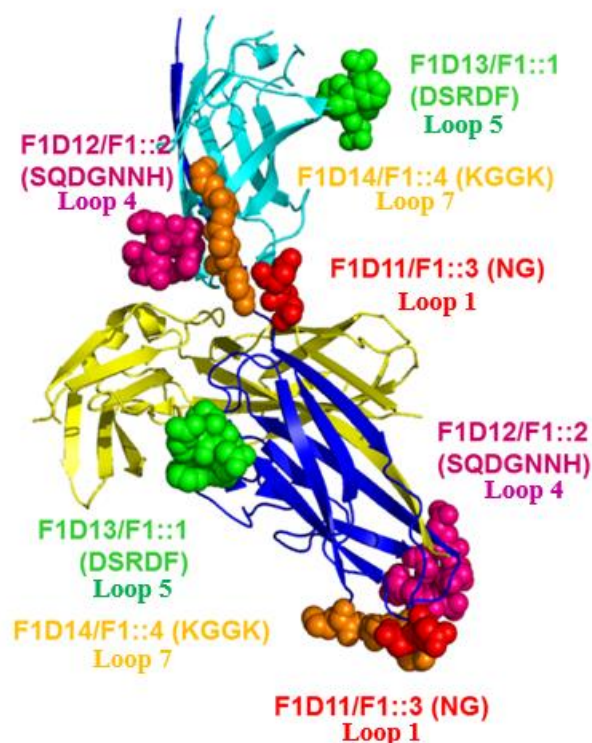


Figure 5.3: Caf1₂:Caf1M complex (PDB: 1Z9S, 2.2 Å) showing surface loops tested for epitope replacement. Yellow, Caf1M; blue, bottom subunit; cyan, top subunit. Sequences in brackets indicate residues replaced by glycine (F1D11, F1D12, F1D3, F1D14) and HIV gp41 core epitope, ELDKWA (F1::1, F1::2, F1::3, F1::4).

5.3 Generation and phenotype of constructs with glycine and charge epitope inserts

To test the efficiency of surface assembly of F1 with each of the loop mutations, DNA fragments of *caf1* encoding loop changes were amplified from the eight respective pFM1 constructs (Appendix 3) and inserted into pACYCF1, by Infusion technology. Figure 5.4 (A) shows the profile of each plasmid preparation used as template for amplification of inserts. There was some concern about the appearance of two bands of supercoiled pACYCF1 (Figure 5.4 (A), lane A). Therefore, two preparations of pACYCF1 (5) and (6) were obtained, transformed into *E.coli* DH5 α and TOP10. Plasmid was prepared from 4 transformants of each. All were the same size (Figure 5.4 (B)). Therefore, all were stored for subsequent use.

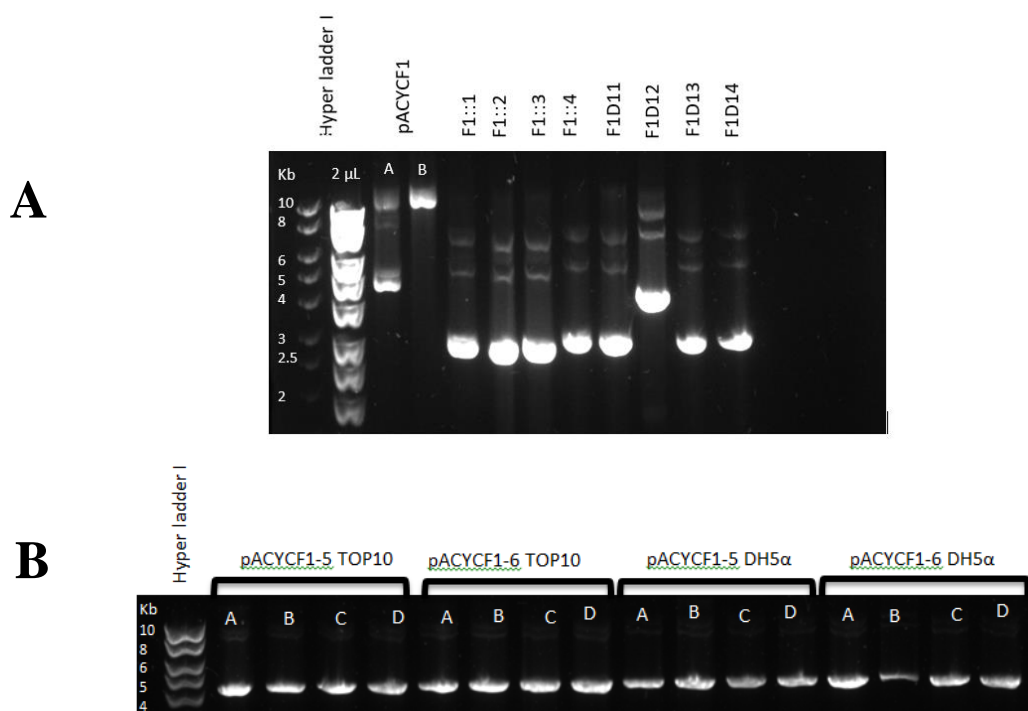


Figure 5.4: Plasmid source of mutated loop sequences
 (A) Electrophoresis gel of undigested plasmids used as templates for constructing loop mutants. The size of pFM1 is 5703bp and the size of pFMA1 used for F1D12 is 8224bp (there was no stock from pFM1 plasmid for F1D12). pACYCF1 A and B were from two individual transformants. (B) Electrophoresis gel displaying multiple preparation of pACYCF1 plasmid extraction of vector templates for constructing loop mutants. No variation in plasmids from sixteen different transformants.

Table 5.1: Caf1 loop mutation in pACYCF1.

Loop mutant	Caf1 residues replaced	Inserted residue	Insert source
pACYC-F1D11	D33N34	GG	pFM1-F1D11 (R665 clone 4)
pACYC-F1D12	S76QDGNNH82	GGGG	pFMA1-F1D12 (R784 clone 1)
pACYC-F1D13	D92SRDF96	GGGG	pFM1-F1D13 (R669 clone A)
pACYC-F1D14	K129GGK132	DGGG	pFM1-F1D14 (R902 clone 2)
pACYC-F1::1	D92SRDF96	ELDKWA	pFM1-F1::1 (R899 clone 1)
pACYC-F1::2	S76QDGNNH82	ELDKWA	pFM1-F1::2 (R900 clone 1)
pACYC-F1::3	D33N34	ELDKWA	pFM1-F1::3 (R901 clone 2)
pACYC-F1::4	K129GGK132	ELDKWA	pFM1-F1::4 (R902 clone 2)

Table 5.1 lists details of the Caf1 loop mutants constructed, pACYC-F1D11, -F1D12, -F1D13, -F1D14, -F1::1, -F1::2, -F2::3 and -F1::4 and the pFM1 constructs

used as source of mutations. Figure 5.5 overviews the strategy. Because inserts with mutations were amplified from *caf1* and Infused back into *caf1* within the pACYCF1 vector, amplification primers were identical to the vector (Table 2.3). The same primers F1for11 and F1rev11 were used to amplify fragment of *caf1* for most constructs. Only –F1D14 and –F1::4 amplified a different *caf1* fragment using primers F1D14insertFOR-2 and F1D14insertREV-2.

Amplification of inserts and linearization of vector was done using CloneAmp HiFi premix. Different annealing temperatures were used to generate linearized pACYCF1 and at the same time delete the relevant region of *caf1*. For F1D11, F1D12, F1D13 F1::1, F1::2 and F1::3 annealing temperatures of 55.4°C, 57.7°C and 59.7°C were used and for F1D14 and F1::4 temperatures of 57.2°C and 59.2°C were used to generate linearized pACYCF1. For amplification of inserts, three different temperatures (55.4°C, 57.9°C and 59.9°C) were used to amplify F1D13 to optimise PCR conditions. After verifying the PCR product by gel electrophoresis, 57.9°C was selected as the optimum temperature to amplify F1D11, F1D12, F1::1, F1::2 and F1::3 inserts. Again, the annealing temperatures of 57.2°C and 59.2°C were used to amplify F1D14 and F1::4, to optimise the PCR conditions as these inserts used different primers. PCR products of inserts and linearized vector were DpnI treated, purified and analysed by gel electrophoresis to check the quality and quantity of the purified PCR products (refer Figure 5.6). Then the inserts were InFused into linearized pACYCF1 and transformed into Stellar competent cells (Clontech). The correct constructs were identified by gel electrophoresis and sequencing. All constructs were confirmed to carry the required loop mutations (Appendix 5-16). Correct constructs were then tested for F1 expression.

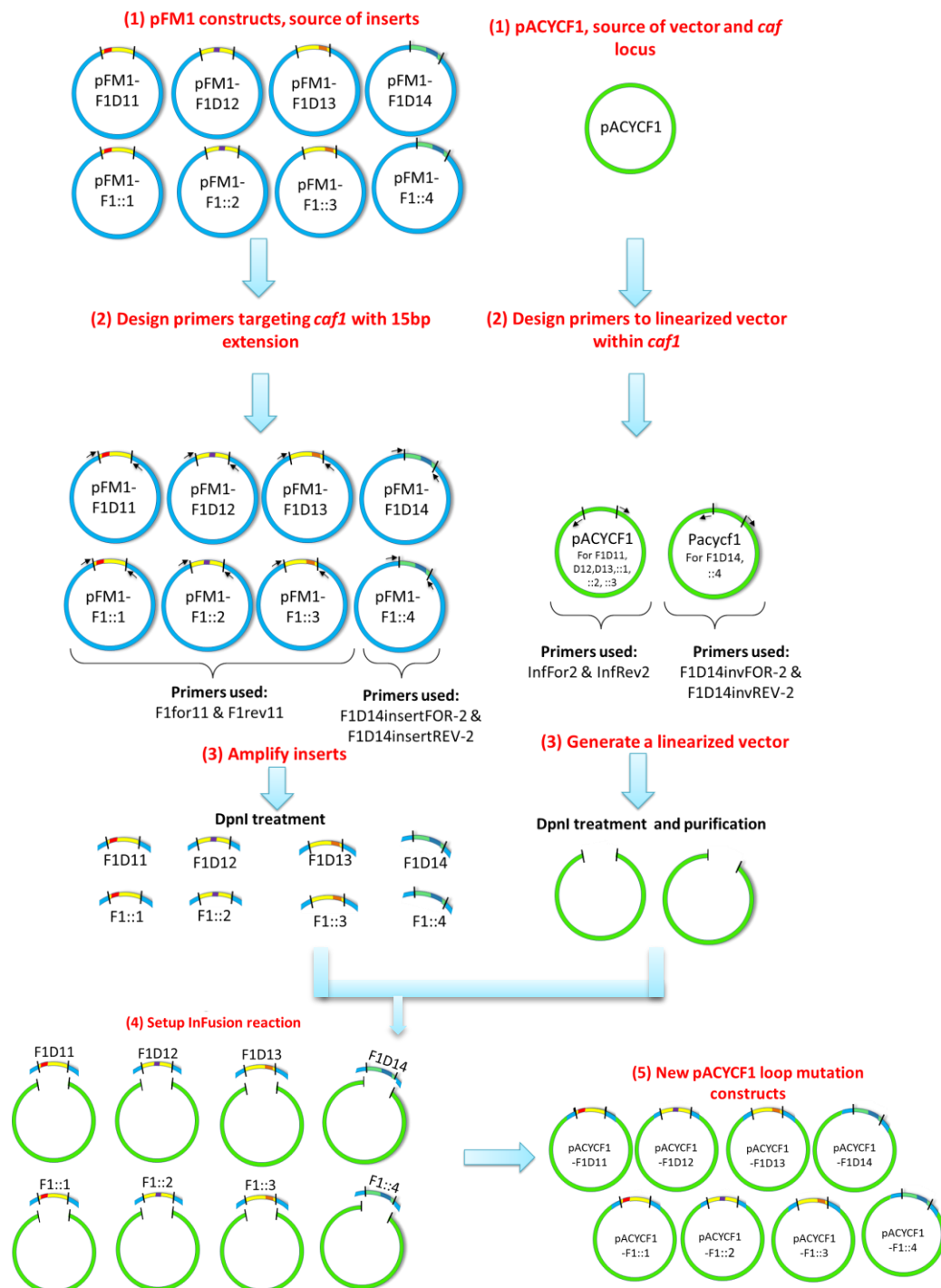


Figure 5.5: Workflow of In-fusion HD protocol.

(1) Appropriately mutated pFM1 constructs and pACYCF1 were obtained as source of inserts and vector, respectively (2) Four sets of primers were designed, two sets used to amplify vector and inserts for F1D11, F1D12, F1D13, F1::1, F1::2 and F1::3. Another two sets were used to amplify vector and inserts for F1D14 and F1::4. (3) Inserts were generated by standard PCR and vectors were linearised by inverse PCR. (4) InFusion reactions were setup to generate all eight constructs (5). Yellow and dark green indicate *caf* gene; bright green, pACYCF1; red, purple, orange and blue, inserts within *caf1*.

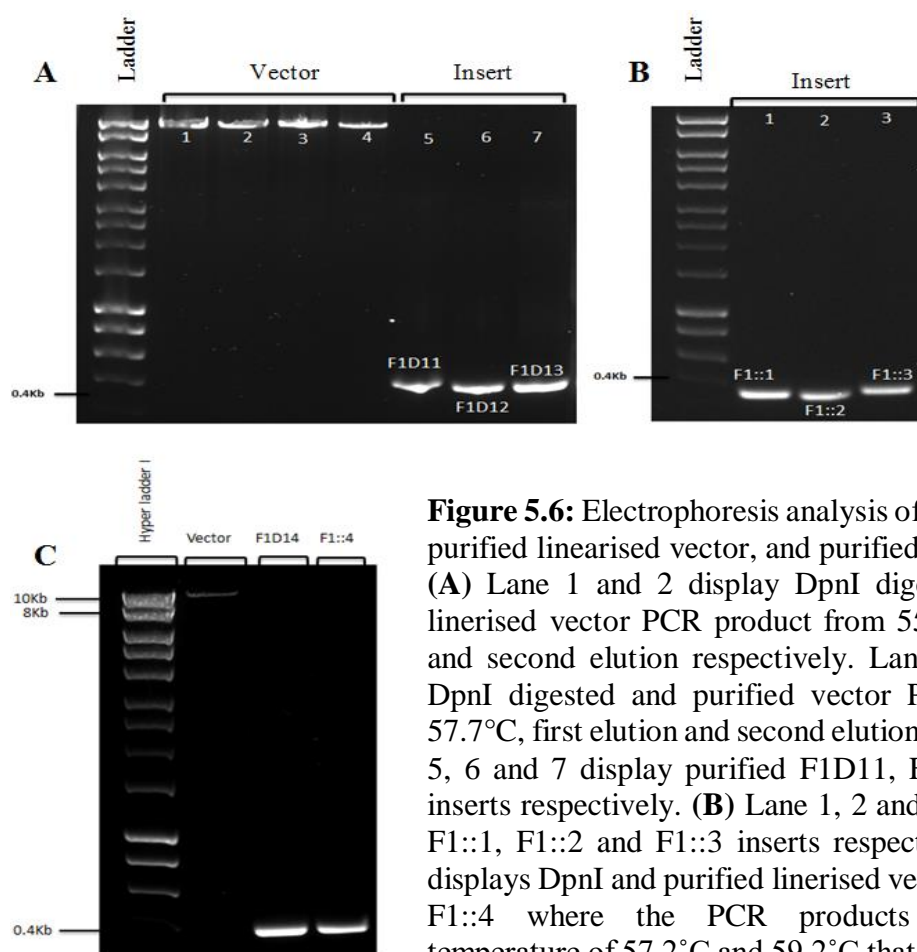


Figure 5.6: Electrophoresis analysis of DpnI digested and purified linearised vector, and purified inserts.

(A) Lane 1 and 2 display DpnI digested and purified linearised vector PCR product from 55.4°C, first elution and second elution respectively. Lane 3 and 4 display DpnI digested and purified vector PCR product from 57.7°C, first elution and second elution respectively. Lane 5, 6 and 7 display purified F1D11, F1D12 and F1D13 inserts respectively. (B) Lane 1, 2 and 3 display purified F1::1, F1::2 and F1::3 inserts respectively. (C) Lane 1 displays DpnI and purified linearised vector for F1D14 and F1::4 where the PCR products from annealing temperature of 57.2°C and 59.2°C that were pooled. Lane 2 and 3 display purified F1D14 and F1::4 inserts respectively.

5.3.1 Effect of F1 loop mutations on bacterial cell phenotype, initial screen

Following construction of all loop mutants, three transformants of each construct, with the correct insert, were screened to monitor the level of F1 expression. Cell pellets were initially screened for the characteristic diffuse white pellet formed on recovery of cells assembling large amounts of F1 (Figure 5.7 (A)). Whole cells were also screened by SDS-PAGE for the prominent band of denatured Caf1 of 15-16 kDa. Transformants carrying plasmid pACYC-F1D11, -F1D13, F1D14 or F1::1 have slightly less F1 compared to wildtype as shown in phenotypic analysis of cell pellets and SDS-PAGE of whole cells. However, transformants carrying the plasmids pACYC-F1D12, F1::2, F1::3 and F1::4 all showed a reduced level of surface F1 (Figure 5.7). To confirm that no additional spontaneous mutation had occurred that could interfere with F1 expression, each of the plasmids resulting in reduced F1 expression was sent for whole *caf* operon sequencing. Seven sequencing primers were used (refer Appendix 13 (E)) and the sequencing results were assembled using SeqMan Pro tool from DNASTar programme. As shown in Appendix 13, the complete *caf* operon in pACYC-F1D12, F1::2, F1::3 and F1::4 constructs were confirmed to have the correct loop mutation and no other spontaneous mutation.

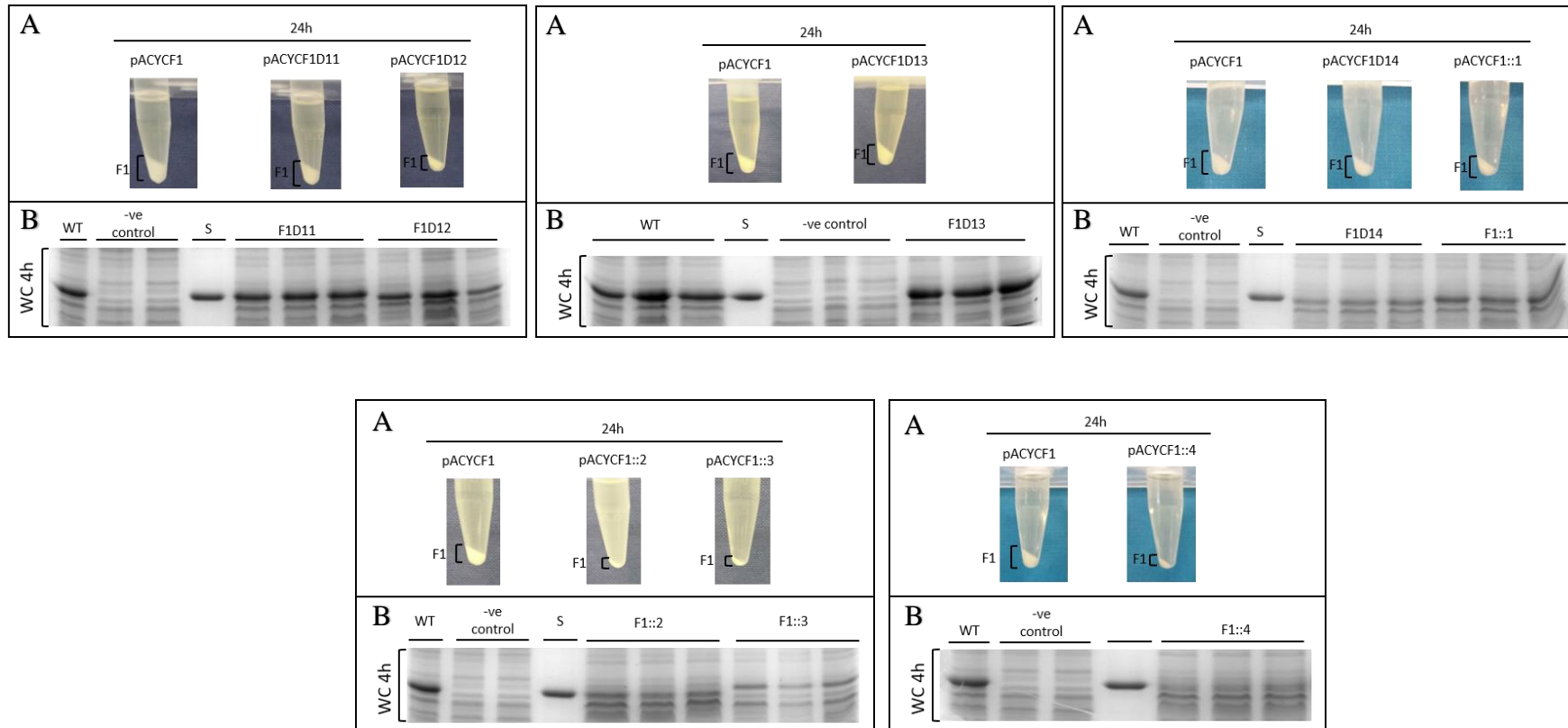


Figure 5.7: Comparison of F1 production from *E. coli* Top10 carrying pACYCF1 encoding different loop mutations.

(A) Visual inspection of cell pellet, after 24 hour thermo-induction at 37°C. The culture (1 ml) was centrifuged at 13000 rpm for 20 minutes. F1 production (WT) is visible as an extensive white fluffy pellet. (B) SDS-PAGE analysis of whole cells (WC). 4 OD cells after 4 hour thermoinduction were recovered, resuspended in 100µl PBS and samples denatured in sample buffer prior to electrophoresis. Replicates are 3 different transformants, each confirmed for the correct insert. WT, pACYCF1; control, pACYCduet1; S, F1 standard.

5.3.2 Impact of loop mutations on surface assembly of Caf1

Following the preliminary screen of F1 loop mutants, confirmed plasmid from one transformant for each construct was transformed into *E. coli* BW25113 to assess the impact on surface assembly. This strain was used as it does not carry a mutation in the *recA* gene and high levels of F1 can be recovered. A transformant colony from each construct was induced for 4 hours, and then heat extracted F1 (65°C/1 hour) was analysed by SDS-PAGE (Figure 5.8). F1 was present in all constructs with loops replaced with glycine. *E.coli*/pACYC-F1D13 produced approximately the same level of F1 as wild type. *E.coli*/pACYCF1-D11 produced slightly less than wildtype and pACYCF1-D12 and -D14 produced the lowest level although, still at a good level of F1 production. In constructs where loops were replaced with ELDKWA, pACYC-F1::1 led to accumulation of reasonably high levels of F1, roughly comparable to that produced from pACYC-F1D12, but much lower than the Gly replacement in the same position (F1D13). Transformants carrying pACYC-F1::2 had a very lower level of F1 and those carrying pACYC-F1::3 or pACYC-F1::4 had a low level of F1. It is also clear that the Caf1 protein produced from pACYC-F1::3 is larger than the other loop mutants. This correlates with the respective changes in the size of the loop, larger for F1::3 compared to wildtype. Efficient polymerisation of all Gly replacement mutants and that of F1::1 (Figure 5.8 (B)) was evident from the recovery of Caf1 as very high molecular weight (HMW) polymer in the stacking gel of unheated extracts (Yu et al., 2009). No HMW polymer could be seen in CB stained gels of samples F1::3, although bands corresponding to oligomers of Caf1 were clearly visible, suggesting problems with polymerisation or stability. Similar putative oligomers and denatured Caf1 were

observed with F1D11 (same position) and F1D12. In fact only F1D13 appeared to behave like wild-type F1 with little visible denatured Caf1 in unheated preparations (Figure 5.8 (B)).

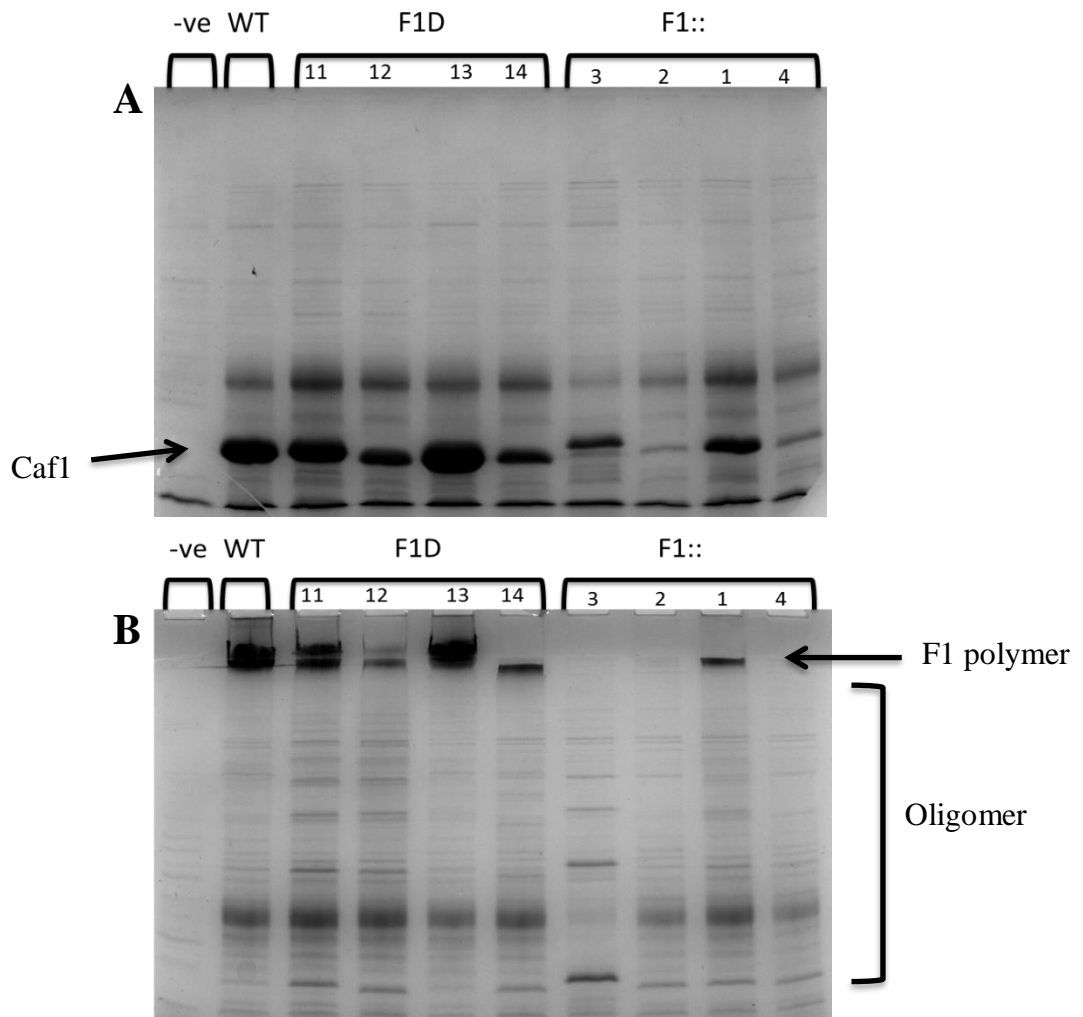


Figure 5.8: SDS-PAGE of surface F1 with loop replacements. (A) F1 extracted from whole cells at 65°C for 1 hour (B) F1 polymer (stack gel) and Caf1 oligomer (separating gel) in unheated samples. Surface extracts were prepared from 3.5 OD cells after 4 hour thermoinduction recovered and resuspended in 100μl PBS.

5.3.3 Impact of loop mutations on stability of the F1 polymer

From phenotypic analysis of loop mutants, it is clear that five constructs (pACYC-F1D11, -F1D12, -F1D13, -F1D14 and -F1::1) assemble Caf1 into high molecular weight polymers, but that some may either polymerise less efficiently or form less stable polymers. To assess the impact of each mutation on stability of these F1 polymers, surface extracted F1 was subjected to heat denaturation in SDS from 65°C to 95°C (see Figure 5.9 for detail). As shown in Figure 5.9 (A), wildtype F1 was stable in SDS up to between 70°C-75°C. At 75°C both oligomers and denatured Caf1 monomer were visible. At 80°C and 85°C, F1 appears to be completely denatured to the monomer. Thus, to ensure complete denaturation of F1 polymer in SDS, temperatures of more than 85°C for 10 minutes should be used to for denaturation.

From Figure 5.9 (B), it is evident that surface F1 of the F1D11 and F1::1 constructs were less stable than wildtype. Both were completely denatured at 75°C. This was confirmed by western blotting (Figure 5.10) where, only monomer and minor band of dimer (30 kDa) was present for F1D11 or dimer (30 kDa) and trimer (40 kDa) with F1::1. The constructs F1D12, F1D13 and F1D14 were all more stable. Coomassie blue stained oligomers can clearly be observed after heating at 75°C, and fully denatured at 85°C. This was confirmed by western blotting for all three constructs (Figure 5.10). However, while the F1D13 construct behaved in a similar manner to wildtype with primarily high molecular weight polymer and little monomer after heating at 75°C, F1D12 and F1D14 were clearly less stable. Both of these constructs, contained a higher proportion of denatured Caf1 after heating at 75°C for 10 minutes, as well as prominent oligomers.

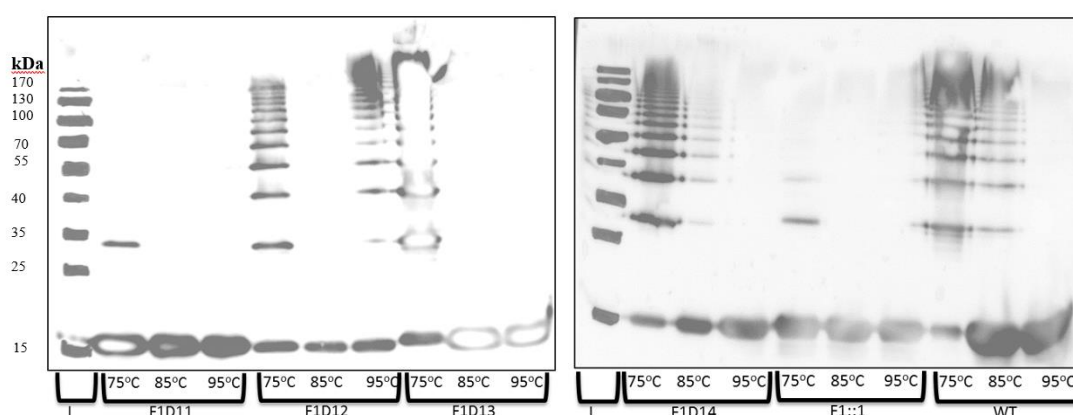
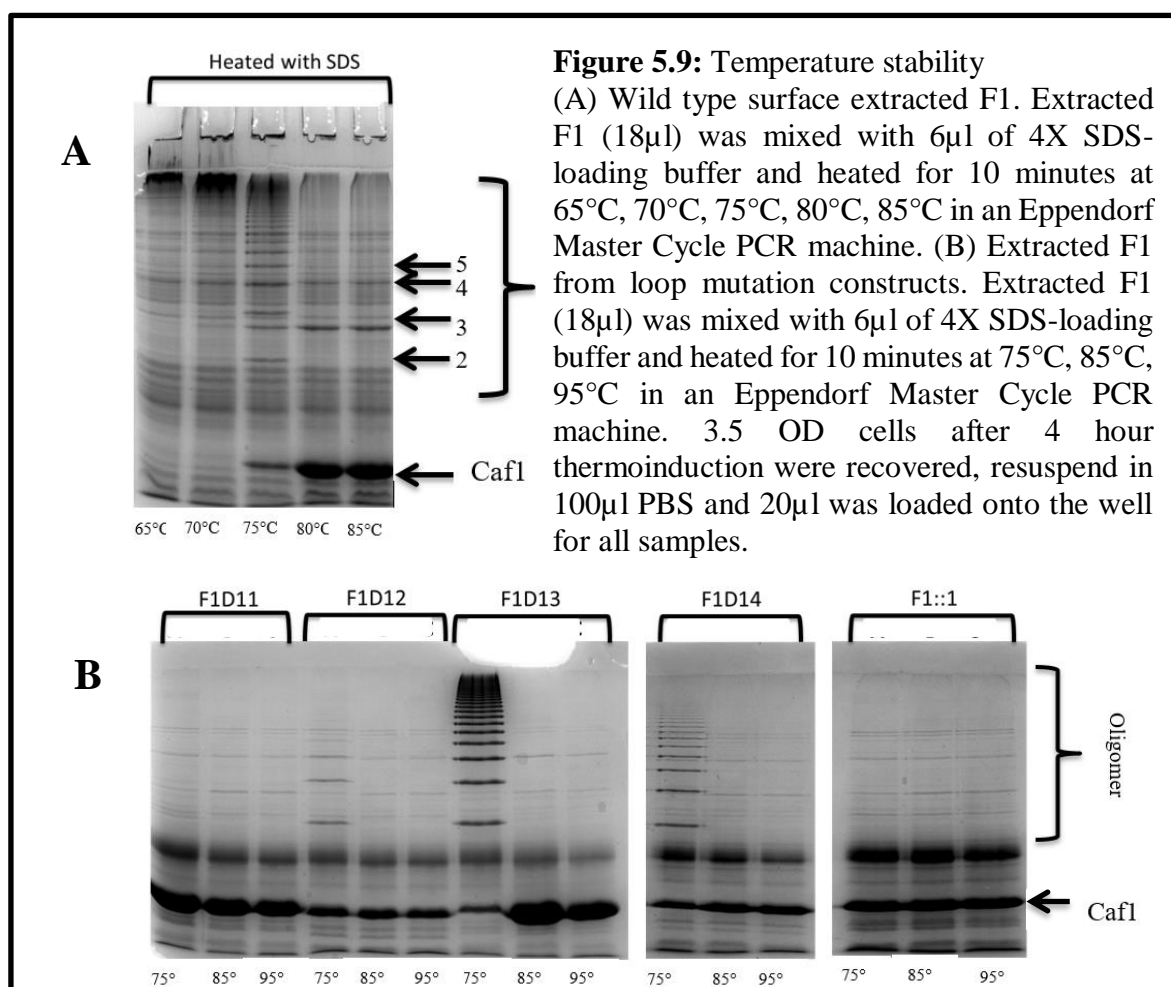


Figure 5.10: Western blot of temperature stability of extracted F1 from loop mutation constructs.

Samples prepared as for Figure 5.9. 20μl was loaded onto the well. L indicates ladder (Thermo Scientific pre-stained protein ladder - 26616). *Due to high level of wildtype and F1D13 polymer, spill over was detected by western blot in neighbouring lanes.*

5.3.4 Quantitative comparison of mutated F1 polymers on the cell surface

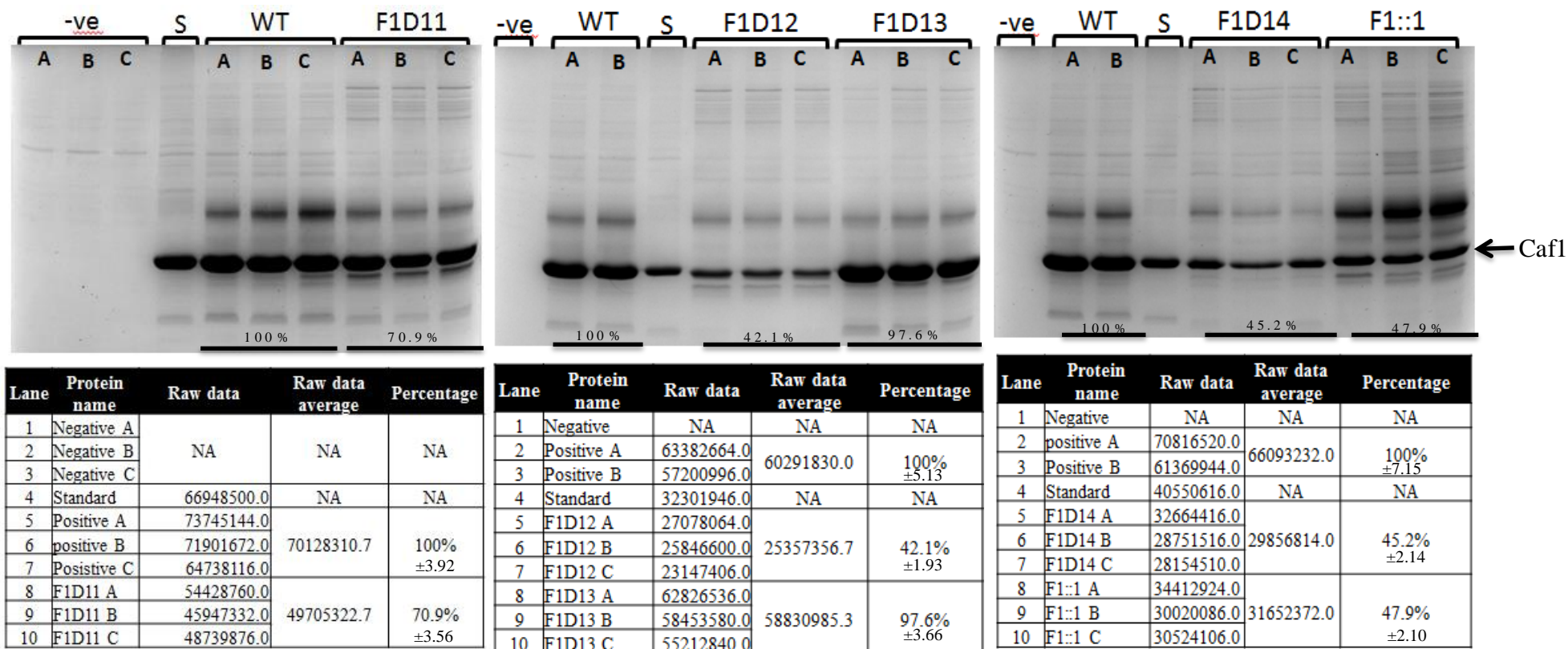
To quantitate differences between the constructs, three fresh transformants of each mutated plasmid in *E. coli* BW25113 were cultured and thermoinduced for 4 hour at 37°C. Surface F1 was then, heat extracted at 65°C for 1 hour and analysed by SDS-PAGE. GeneTool programme was used for quantitation of the Caf1 band by averaging the raw data of the triplicate samples and subtracting the background. The average for wildtype F1 was assigned as 100 % and the efficiency of surface assembly of each mutant expressed as a percent of wild-type (Figure 5.11 and Figure 5.12).

Results confirmed the preliminary phenotype shown in Figure 5.7. Transformants carrying insertions in loop 1, pACYC-F1D11 (Gly insertion) and pACYC-F1::3 (ELDKWA insertion), produced surface F1 but at slightly lower levels with an average of $70.9\% \pm 3.56$ and $30.2\% \pm 4.18$ surface F1, respectively, compared to the wildtype surface F1 (100%). Analysis of unheated preparations of both constructs revealed accumulation of oligomers in addition to some HMW polymer (Figure 5.8 and Figure 5.10). Hence, insertions within loop 1 are likely to interfere with the rate of polymerisation and this is exaggerated with the larger charged insertion.

Quantitation of extracted F1 from cells carrying pACYC-F1D12 (Gly insertion) and pACYC-F1::2 (ELDKWA insertion), insertions in loop 4, revealed assembly of only $42.1\% \pm 1.93$ and $6.9\% \pm 1.45$ surface F1, respectively, compared to wildtype. Cells carrying pACYC-F1D14 (Gly insertion) and pACYC-F1::4 (ELDKWA insertion), insertions in loop 7, showed an equally reduced level of polymer formation with values of $45.2\% \pm 2.14$ and $8.9\% \pm 1.15$ respectively, compared to surface F1 from the wild type.

Again, the unheated surface F1 (refer to Figure 5.8 (B)) from both constructs revealed oligomers consistent with interference with the assembly of Caf1.

In contrast, insertions in loop 5, had much less impact on surface assembly. Cells carrying pACYC-F1D13 (Gly insertion) or pACYC-F1::1 (ELDKWA insertion) produced $97.6\% \pm 3.66$ and $47.9\% \pm 2.14$ of surface F1, respectively, compared to wildtype. Moreover, F1D13 polymer was as stable as wild type F1. Despite the lower level of high molecular weight polymer produced from construct pACYCF1::1, this construct produced the most F1 of all the constructs carrying a charged epitope insertion. From this data, loop 5 would appear to be an optimum site for insertion of foreign epitope.



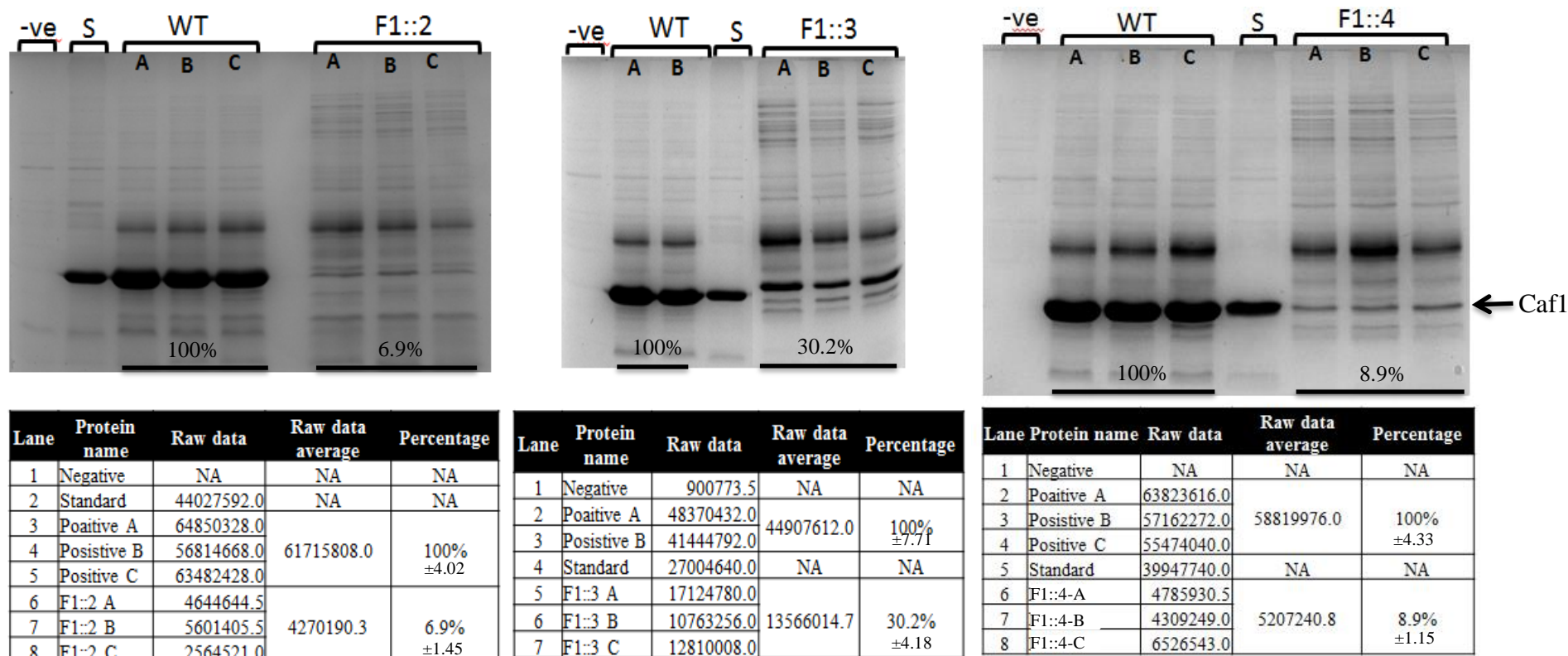


Figure 5.12: Quantitation of surface F1 extracted from *E. coli* BW25113 expressing F1 with loop mutations (pACYC-F1::2, -F1::3 and -F1::4). Top images, SDS-PAGE of F1 extracted at 65°C/1 hour in triplicate for each construct as shown. Lower images, quantitation of imaged F1 band using Syngene programme. WT indicates wildtype F1 assigned as 100% for each gel. S indicates F1 standard (5µl applied).

5.4 *In silico* analysis of F1 loop insertions

To assess the reason for poor F1 fibre assembly in some construct, the Caf translocon accommodating subunits with loop replacements were modeled. IntFOLD2 was used to model the fiber inserted subunit where the best template selected by this server was 1Z9S:C (Caf1M:Caf1:Caf1 complex). I-TASSER was used to model the bottom subunit with 1Z9S:B assigned as template. There was a need to model the modified bottom and top subunits as the cleft in the top subunit is filled with the N-terminal Gd strand from the bottom subunit; and the cleft in the bottom subunit, is filled with the chaperone G1 donor strand from chaperone. Each FASTA sequence used as input in I-TASSER can be seen in Appendix 4.

Table 5.2 shows that all 16 Caf1 subunits with loop mutations modelled as chaperone bound (IntFOLD2) or as fibre inserted (I-TASSER) have a very high confidence value indicating that they were correctly modelled. Small loops are likely to be unstructured but to assess any conformational differences at the regions with insertions, the Caf1 models (from IntFOLD2) were superimposed with Caf1 structure (1Z9S:B) using PyMOL (Figure 5.13). No conformational changes were predicted in Caf-F1D11, F1D14, F1::2 and F1::4. The size of loop was reduced in F1D12 where residues in loop 4 were reduced from seven to four amino acids and in F1::3 loop 1 was converted from two to six amino acids, making a bigger loop. In F1D13, where residues in loop 5 were reduced from five to four amino acids. This construct was modeled with a shorter loop and missing a β -strand, yet produced the highest level of surface F1 and the most stable modified F1 polymers. F1::1 where residues in loop 5 were converted from five to six amino acids, possessed a bigger loop missing the α -helix.

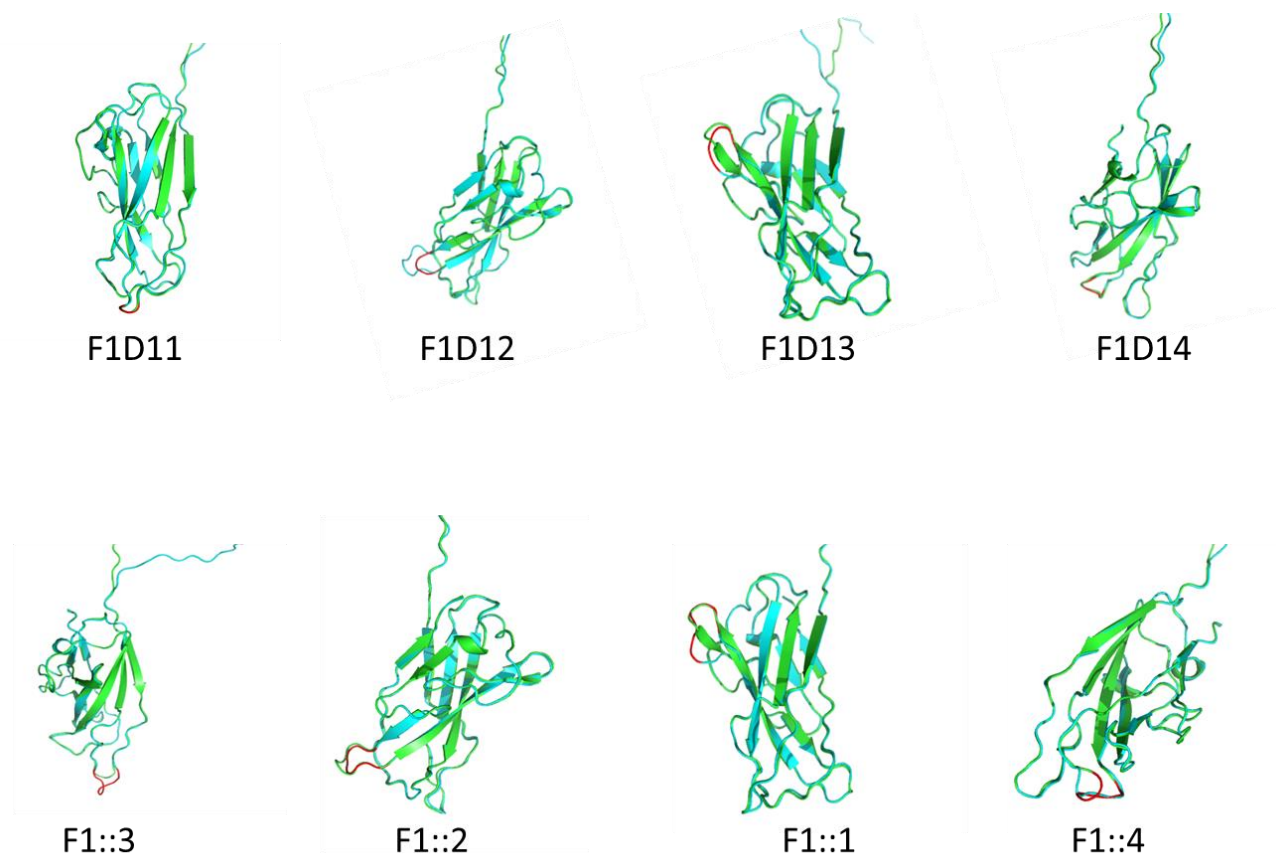


Figure 5.13: Superimposed Caf1 mutated model with Caf1 wild type structure to compare the size of insertion in each loops. Coloured in red is the replacement either glycine or ELDKWA; cyan, mutated model; and green, Caf1 WT structure (1Z9S:B, chaperone bound subunit).

Caf1 with loop mutation model	Global model quality score (chaperone bound subunit, bottom subunit)	C-score (N-terminal extension inserted subunit, top subunit)
F1D11	0.9511	1.55
F1D12	0.9416	1.53
F1D13	0.9500	1.49
F1D14	0.9586	1.57
F1::1	0.9507	1.52
F1::2	0.9492	1.50
F1::3	0.9324	1.40
F1::4	0.9485	1.58

Table 5.2: Quality and confidence level of each mutated Caf1 model.

The confidence value of IntFOLD2, Global model quality score (range 0-1, value >0.4 indicates a more complete and confident model) and I-TASSER, C-score (range between -5 and 2, higher value signifies high confidence model).

Models of the mutated subunits (fibre inserted and chaperone bound for each of the 8 mutated subunits) were then assembled with the Caf1A model to form the 8 different translocons; thus, permitting analysis of required interactions or clashes between the mutated residues and the usher that might affect F1 assembly. To assemble each translocon, TMalign was used to superpose Caf1A on FimD, 3RFZ chain B (TM-score=0.91412). Caf1M (1Z9S chain A) was then superposed with FimC (3RFZ chain C) to gain coordinates for Caf1M relative to FimC (TM-score=1). In PyMol, the Caf1A model was aligned with the TMalign file of Caf1A based on FimD. Then, Caf1M:Caf1 complex (1Z9S:A and B) was aligned with the TMalign file of Caf1M based on FimC. Subsequently, the bottom Caf1 mutated model was aligned to the Caf1M:Caf1 bottom complex to assemble the bottom Caf1 at this position. The process was repeated for the top Caf1 mutated model. The Caf1M structure was aligned to Caf1M:Caf1 complex to complete the translocon. Finally, The Gd strand from 1Z9S chain B (residue 1 to 20) was loaded into PyMOL and aligned to the bottom subunit model to replace the disordered N-terminal Gd extension. The entire process was repeated to assemble all

eight translocon models each with two mutated subunits. All are available in the attached CD (see Appendix 14 for file details).

Each translocon was checked for any change in interaction between the modified Caf1 subunit and usher. For Caf-F1D11, there is no interaction between the inserted glycines with any residues in the translocon Figure 5.14 (Caf1-F1D11) (A)(iii). However, with the longer charged insert in the same loop (Caf1-F1::3), Gln33 and Leu34 of the inserted epitope clash with Arg18 and Pro16 in the Gd strand of the incoming, chaperone-bound Caf1 (Figure 5.14 (Caf1-F1::3)(B) (iii)). This would be consistent with the observed interference in polymerisation of Caf1 carrying the inserted epitope. This was validated by the low level of F1 produced (refer to section 5.3.4) where pACYCF1::3 produced $30.2\% \pm 4.18$ surface F1. The inability to polymerise efficiently had also been observed earlier when Caf1-F1::3 was expressed from pFM1 and oligomers of only 1-8 subunits were recovered (pFM1 encodes chaperone and subunit only and results in polymerisation in the periplasm). This was in contrast to wildtype and Caf1-F1D11, which assembled high molecular weight polymer (refer Appendix 3). Thus, changes to loop 1 are likely to interfere with subunit: subunit interaction, decreasing efficiency of polymerisation.

For Caf-F1D12 and Caf-F1::2 (loop 4 inserts), two interactions are lost, Tyr266(plug) with Asn81(subunit) and Thr267(plug) with Gly131(subunit) (Figure 5.11 and Figure 5.12). In addition, for Caf1-F1::2, the inserted Trp80 within the epitope clashed with Tyr266 in the plug. A comparison of both modelled translocons in Figure 5.14, would suggest that there is nothing hindering Caf1-F1D12 from passing through the Caf1A barrel while entrance of Caf1-F1::2 into the usher barrel would be blocked

by the additional epitope. This would explain the exceptionally low level of surface F1::2 recovered ($6.9\% \pm 1.45$ compared to wildtype). The poor recovery of F1D12 (45 % of wild type) might be attributed to loss of required interactions with usher. This is supported by the fact that when expressed from pFM1, wildtype levels of F1D12 polymer are produced (Appendix 3) while less surface F1 is produced from pACYCF1D12, it still forms relatively stable polymer. The Asn80 residue is conserved in multiple sequence alignment of Caf1 and subunits from *E. coli* SE11. It was also lost with insertions in loop 4. Thus, possession of either Asn80 or Asn 81 might be critical in translocation of Caf1.

Figure 5.14 identifies an interaction of 3.5\AA in Caf1:F1D13 between Gly92 (inserted glycine in subunit) and Gln169 (Caf1A barrel). When checked in the wild type translocon, Gln169 does not interact with any residue. However, there is an interaction of 3.6\AA in the wild type translocon between Ser93 (subunit) and Gln167 (Caf1A barrel). Based on the quantitation, pACYCF1D13 construct produced essentially wildtype levels of assembled polymer ($97.6\% \pm 3.66$). Thus, if the Ser93-Gln167 interaction is important for assembly of F1 the new Gly92 - Gln169 interaction may replace this requirement to maintain efficient export of F1 fibre to cell surface. The stability of the F1 fibre from this construct was also shown to be as stable as wildtype. With the changed epitope in this site, pACYCF1::1, there is a clash between Leu93 (inserted epitope of top subunit) and Gln169 (Caf1A barrel) and an interaction of 3\AA between Asp94 (inserted epitope of top subunit) with Ser182 (Caf1A barrel) (refer to Figure 5.14). However, the interaction between Ser93 and Gln167 that occurred in the wild

type was also lost with this epitope insertion. The loss of interaction and clash in this construct could explain the reduced level of surface F1 expressed of only $47.9\% \pm 2.14$.

Figure 5.14 also shows an interaction of 3\AA of Caf1-F1D14 Gly131 (subunit) with Thr267 (Caf1A plug) and interaction of 2.4\AA between Asn81 (subunit) with Tyr266 (plug). This latter interaction is the same as seen in wildtype. However, in Caf1-F1::4, Asn80 and Asn81 interact with Tyr266 by 2.1\AA and 3.1\AA , respectively. In the wildtype translocon, Asn81 interacts with Tyr266 by 3\AA and Gly131 interacts with Thr267 by 2.6\AA . In Caf1-F1D14, both interactions remained and this predicts that the assembly and export of F1 will not be abolished. This was confirmed by quantitation where the level of surface F1 was $45.2\% \pm 2.14$ and shared typical wildtype heat-stability properties: after heating at 75°C in SDS, partially denatured polymer was observed as a ladder of oligomers (refer to Figure 5.9). The interactions that are lost in F1::4 remain in F1D14 which would correlate with the much higher level of exported F1 with this construct, where only $8.9\% \pm 1.15$ exported outside of the cells carrying pACYCF1::4. The very low level of F1 from Caf-F1::4 transformants can be explained as the interaction between the subunit (Gly131) and plug (Thr267) was lost (see Figure 5.14) and interfering with the opening of the pore.

In conclusion, transformants carrying pACYC-F1D11 and -F1D13 give a very good level of surface F1. This indicates that changes in these regions do not interfere with the assembly or export of F1. However, for loop mutations inserted with ELDKWA, only pACYC-F1::1 gives a reasonable level of surface F1. Therefore, loop 5 would be a good target for future loop replacements. Also, it is possible that genetic manipulation of pACYC-F1::1 could be performed to enhance export.

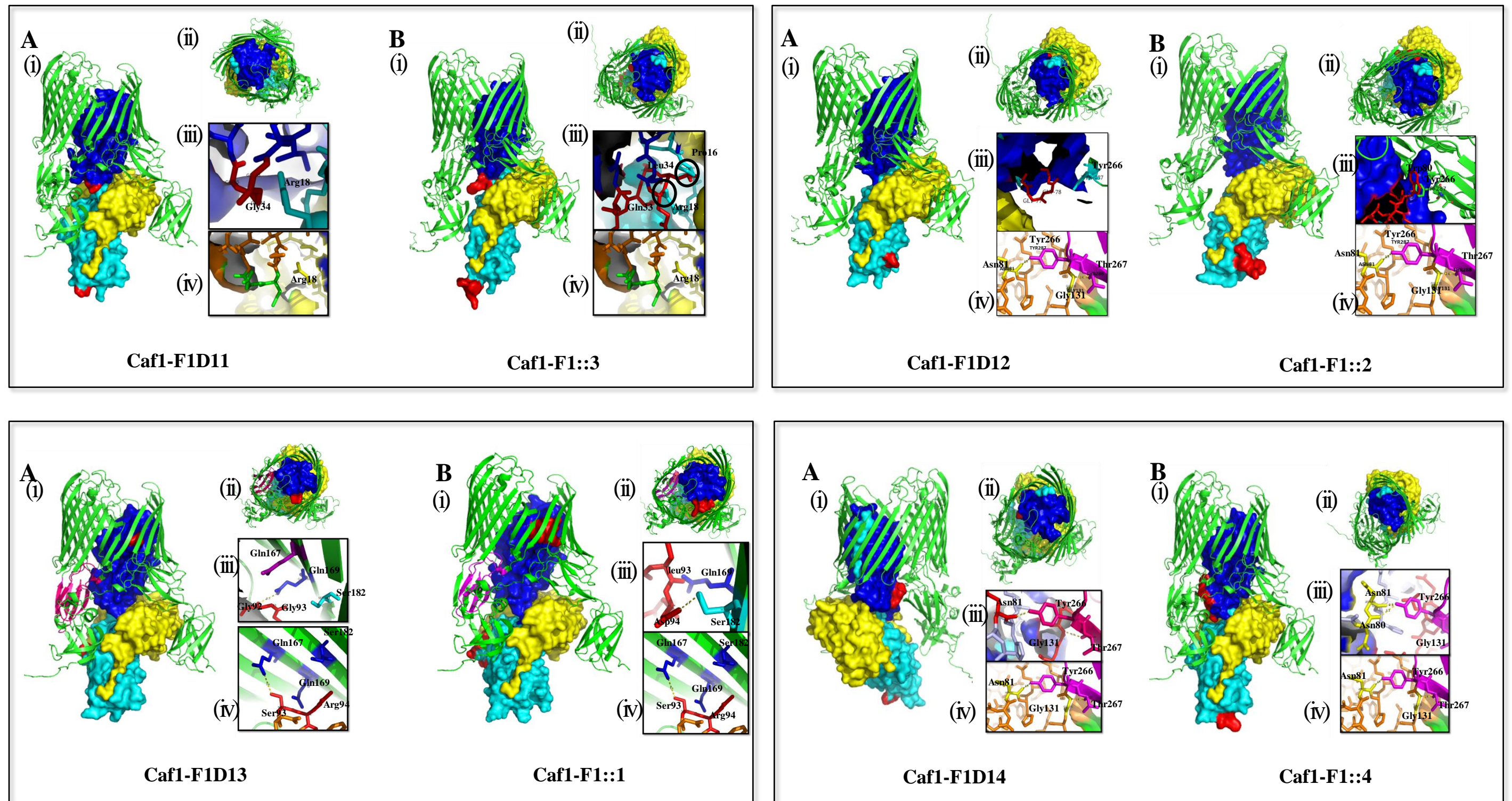


Figure 5.14: Transbicon of loop mutation.

(A) Transbicon with CafI changed to gly. (B) Transbicon with CafI changed to charged epitope, ELDKWA. (i) Model of whole transbicon. Coloured in blue is top CafI, cyan is bottom CafI, yellow is CafIM, green is CafIA and red is charged residues. (ii) View of the transbicon from top. (iii) Interaction or clashes between residues in transbicon with mutated CafI. (iv) Corresponding interaction or clashes between wild type protein in transbicon. Caf1-F1D11, no additional interaction or clash. Caf1-F1::3; circled in black, clashes between Gln33 (inserted epitope in the top CafI) with Arg18 (Cd strand of the bottom subunit), Leu34 (inserted epitope in the top subunit) with Pro16 (Cd strand of the bottom subunit). Caf1-F1D12, loss an interaction of Arg81 with Tyr287. Caf1-F1::2; clashed between Trp80 (inserted epitope) clashed with Tyr287 (plug). Caf1D13, new interaction occurred between Gly92 (inserted glycine in subunit) and Gln169 (CafIA barrel) and loss interaction between Ser93 (subunit) with Gln188 (CafIA barrel). Caf1-F1::1, clashed between Leu93 (inserted epitope of top subunit) with Gln169 (CafIA barrel) and loss interaction between Ser93 with Gln188. Caf1-F1D14, no additional interaction or clash. Caf1-F1::4, new interaction between Asn80 with Tyr266 and loss interaction between Gly131 and Thr267.

5.5 Increasing surface production of F1 with loop insertions by site-directed mutagenesis

Following *in silico* analysis of the F1 loop mutant translocons, usher residues identified as clashing with subunit residues were mutated to a smaller residue. In addition, residues that were involved in interactions in the wild-type translocon but lost in the mutated constructs were also tested for their requirement and if appropriate were added back to the modified loop to test for improved export. Table 5.3 lists residues that were mutated and their location in the *caf* operon.

Table 5.3: List of point mutation and insertion to improve F1 export in loop mutation constructs.

	Name of mutation	Location	Mutation position	REMARKS
Mutation to improve export of pACYCD11 and pACYCF1::3	pACYCF1_F1_P16A	Subunit (Gd strand)	P16A	To validate if the residue is important in polymerising Caf1 (clash with added Leu34 in loop mutation F1::3)
	pACYCF1::3_F1_P16A	Subunit (Gd strand)	P16A	To improve polymerisation/export of F1 (export 30.2% in F1::3)
	pACYCF1_F1_R18Q	Subunit (Gd strand)	R18Q	To validate significance in polymerising Caf1 (clash with added Gln33 in loop mutation F1::3)
	pACYCF1::3_F1_R18Q	Subunit (Gd strand)	R18Q	To improve polymerisation/export of F1::3 (export 30.2% in F1::3)
	pACYCF1_F1_R18K	Subunit (Gd strand)	R18K	To validate significance in polymerising Caf1 (clash with added Gln33 in loop mutation F1::3)
	pACYCF1::3_F1_R18K	Subunit (Gd strand)	R18K	To improve polymerisation/export of F1::3 (export 30.2% in F1::3)
Mutation to improve export of pACYCD12 and pACYCF1::2	pACYCF1_A_Y266A	Plug	Y266A	To validate potential influence in opening and closing the pore (clash with added Trp80 in F1::2 - export only 6.9%)
	pACYCF1_A_Y266S	Plug	Y266S	To validate potential influence in opening and closing the pore (clash with added Trp80 in F1::2 - export only 6.9%)
	pACYCF1_F1_N80G	Subunit (loop)	N80G	To validate if this residue/interaction important in binding with Tyr266 in plug/ opening and closing pore in WT
	pACYCF1_F1_N81G	Subunit (loop)	N81G	To validate if this residue/interaction important in binding with Tyr266 in plug/ opening and closing pore in WT

	pACYCF1D12_GGGGNNH	Subunit Loop	Add NH making GGGGNNH insert	To validate if shift GGGG three residues backwards can increase the export (export 6.9%).
	pACYCF1::2_ELDKWANNH	Subunit Loop	Add NNH making ELDKWANNH insert	To validate if shift ELDKWA three residues backwards and avoiding a clash between Tyr266 and added Trp can increase the export (export 42.1%).
	pACYCF1::2_A_ELDKWANH	Subunit Loop	Add NNH making ELDKWANH insert	To validate if shift ELDKWA two residues backwards and avoiding a clash between Tyr266 and added Trp can increase the export (export 6.9%).
Mutation to improve export of pACYCD13 and pACYCF1::1	pACYCF1_A_Q169A	Barrel	Q169A	To validate significance for F1 export (clash with added Leu in loop mutation F1::1)
	pACYCF1_A_Q169G	Barrel	Q169G	To validate significance for F1 export (clash with added Leu in loop mutation F1::1)
	pACYCF1::1_A_Q169A	Barrel	Q169A	To improve export of F1::1 by replacing with smaller residue (export 47.9%)
	pACYCF1::1_A_Q169G	Barrel	Q169G	To improve export of F1::1 by replacing with smaller residue (export 47.9%)
	pACYCF1_A_S182A	Barrel	S182A	To validate significance for F1 export. Ser182 interacts with inserted Asp94 in F1::1
	pACYCF1::1_A_S182A	Barrel	S182A	To validate if removal of Ser203 interaction improves F1::1 export. Ser182 interacts with inserted Asp94 in F1::1
	pACYCF1_A_Q167S	Barrel	Q167S	To validate importance of Gln167 – Ser93 interaction, interaction is lost in F1D13 and F1::1
	pACYCF1_A_Q167A	Barrel	Q167A	To validate importance of Gln167 – Ser93 interaction, interaction is lost in F1D13 and F1::1
Mutation to improve export of pACYCD14 and pACYCF1::4	pACYCF1::4_F1_N81G	Subunit (loop)	N81G	To validate if this residue/interaction important in binding with Tyr266 in plug/ opening & closing pore in loop mutations
	pACYCF1::4_F1_N80G	Subunit (loop)	N80G	To validate if this residue/interaction important in binding with Tyr266 in plug/ opening&closing pore in WT

5.5.1 Attempt to repair clashes from loop 1 in the F1::3 construct

With the pACYCF1::3 construct, only 30.2% of F1::3 was exported and this was attributed to a clash between residues of the inserted charge epitope with Arg18 and Pro16 in the Gd strand from the bottom subunit. The nature of Gd strand residues is critical to subunit polymerisation (Yu et al., 2012; Zavialov et al., 2002). However, Arg18 and Pro16 both face outwards and hence, are less likely to be critical for donor strand exchange. Thus, these residues were mutated in wildtype pACYCF1 to assess any impact on production of wild-type F1. Arg18 was mutated to Lys (smaller residue in the same polar group) and Gln (smaller residue), while Pro16 was mutated to a smaller residue, Ala. Neither of the Arg mutations had a negative impact on assembly of F1, R18K, and R18Q $104.24\% \pm 10.43$ and $117.9\% \pm 15.37$, compared to wild-type, respectively (Figure 5.15). In contrast, a small reduction in production of surface polymer, $77.13\% \pm 6.12$ of wildtype F1, was observed with the Caf1P16A mutation. When the same point mutations were created in pACYCF1::1, there was no marked stimulation in production of surface polymer. The levels of mutated F1 were $34.24\% \pm 2.97$, $19.45\% \pm 2.11$ and $37.26\% \pm 3.38$ compared to WT for F1::3_R18K, F1::3_R18Q and F1::3_P16A, respectively. It can be concluded that none of these 3 additional mutations showed a marked enhanced export of F1::3. The reason why F1::3_R18Q was lower, is not clear, but the background staining suggests cell lysis. Whole operon sequencing would exclude any additional mutation. Despite the lack of impact of additional mutation on export of F1::3, this experiment has identified that the outward facing Gd strand residue Arg18 does not appear to be critical in polymerisation. It is possible that the large change from Pro16 to Ala had a negative effect on the

structure of the Gd strand, thus slightly decreasing the efficiency of polymerisation. Additional mutations would be required to assess if the Pro itself is important.

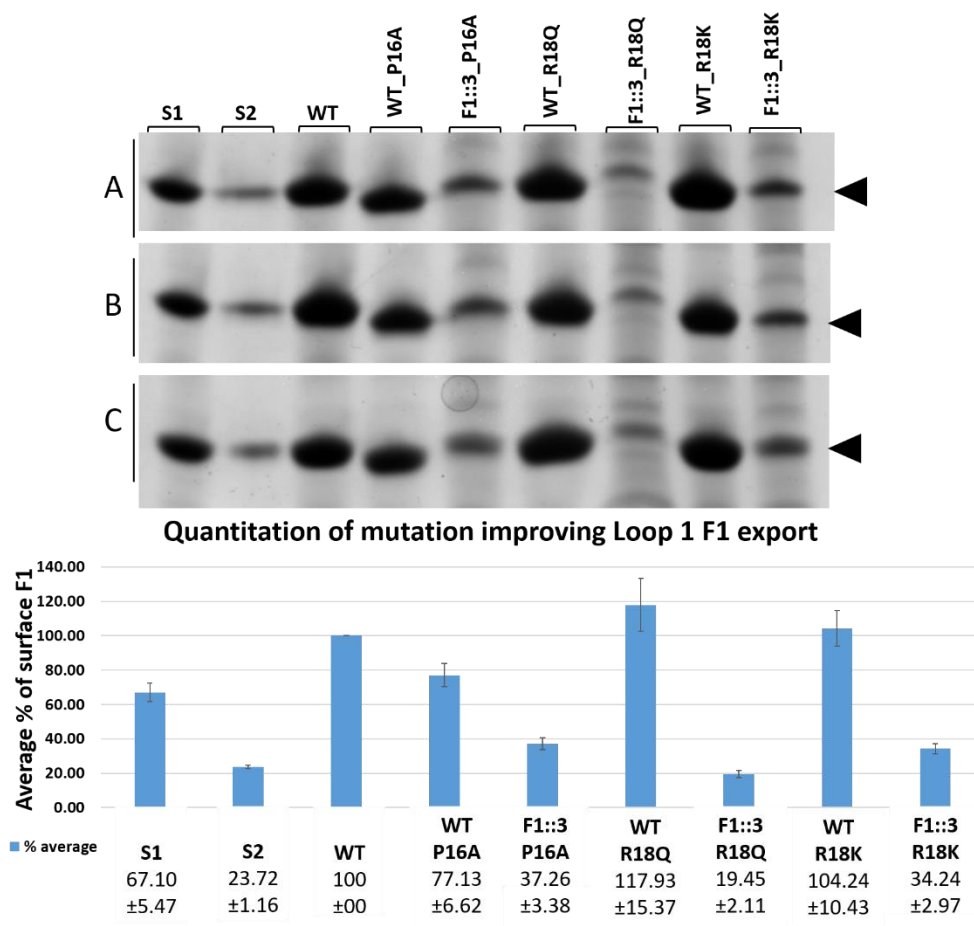


Figure 5.15: SDS-PAGE of extracted surface F1 from *E. coli*/pACYCF1 with point mutations in Caf1A.

Following 4 hour thermoinduction, F1 was extracted at 65°C for 1 h from 3 OD cells in 100µl PBS. Extracted samples (12µl) were mixed with 4µl SDS loading dye, heated at >95°C for 10 min to ensure complete denaturation of F1 fibre. Denatured sample (10µl) was loaded on SDS-PAGE. Following electrophoresis, protein was stained with coomassie blue, imaged and bands of denatured Caf1 monomer (black arrowhead) quantitated. S1 and S2 indicate 2.59µg and 0.863µg, respectively, of the F1 standard applied. A, B and C are biological repeats, 3 separate transformants processed for each construct.

5.5.2 Replacement of required Asn residue in loop 4 of pACYCF1D12 and pACYCF1::2 constructs

The attempt to enhance export of the loop 4 mutant F1::2 focused on loss of the interaction between Tyr266 (plug) and Asn81 (subunit). As discussed in chapter 4, mutation of Caf1A Tyr266 to Ser or Ala, in pACYCF1 resulted in only a slight decrease of surface F1. Therefore, despite the fact that this Tyr-Asn interaction may still be important for efficient F1 assembly, Tyr266 was mutated in the pACYCF1::2 construct to minimise the clash with Trp80 of the inserted sequence in loop 4. However, recovery of surface extracted polymer from cells with either pACYCF1::2_Y266A or pACYCF1::2_Y266S remained extremely low (Figure 5.16). It is possible that this is in part attributable to the importance of Tyr266 in gating the pore (see Figure 4.12 in Section 4.6.1) as well as its interaction with subunit (via Asn81).

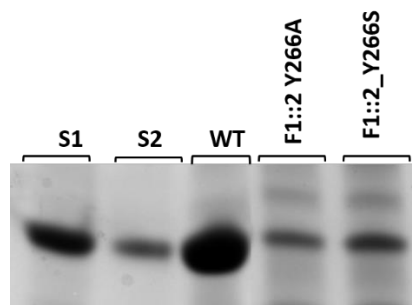


Figure 5.16: SDS-PAGE of surface F1 extracted from *E. coli* BW25113 from construct pACYCF1::2_Y266A and pACYCF1::2_Y266S. 3 OD cells after 4 hour thermoinduction were recovered, extracted and resuspend in 100µl PBS. 10µl of samples were loaded onto each lane.

As removal of the Tyr266 clash did not improve export, focus turned to the Asn residues. Caf1 Asn81 was mutated to the smaller residue from the same polar group, Gly in pACYCF1 (wildtype). There was no impact on export of surface F1. Referring back to the multiple sequence alignment of Caf1 and CU subunits from *E. coli* SE11 (Figure 4.4, section 4.3.2), Asn80 was conserved rather than Asn81. Thus, Caf1Asn80 was also mutated to Gly in pACYCF1 and surface F1-N80G extracted. Surprisingly, the production of F1-N80G in cells carrying pACYCF1_N80G decreased dramatically to $18.4\% \pm 2.18$ of wild-type, suggesting that Asn80 may be essential in the export of F1 to the cell surface (Figure 5.17 and Figure 5.18). In F1D12, the loop size had been reduced from seven residues to four residues and included loss of both Asn residues. Therefore, to test the ability of Asn residue(s), to restore export F1 with modified loop 4, the three residues NNH were reintroduced to pACYCF1D12 and to pACYCF1::2, creating the loop sequences GGGNNH and ELDKWANNH, respectively. Cells carrying pACYCF1D12NNH exhibited a dramatic increase in F1 production ($145.19 \pm 4.92\%$ of wild-type) compared to pACYCF1D12 ($42.71\% \pm 2.25$ of wild-type). However, for cells carrying the pACYCF1::2NNH construct, the level of mutated F1 was slightly reduced to $11.51 \pm 1.53\%$ compared to cells carrying F1::2 ($17.13\% \pm 2.32$). This might be explained by the increased size of the loop interfering with export. Therefore, the construct pACYCF1::2_NH was made by deleting one Asn. This small change increased mutated F1 assembly to $43.63 \pm 2.44\%$ (Figure 5.17 and Figure 5.18). The small decrease in size, correct orientation of Asn, avoiding clashes of other residues with Tyr266 and Thr288 in the plug could each have contributed to improved export. This study provides strong evidence that Asn80 in subunit is essential to export F1 to the cell surface.

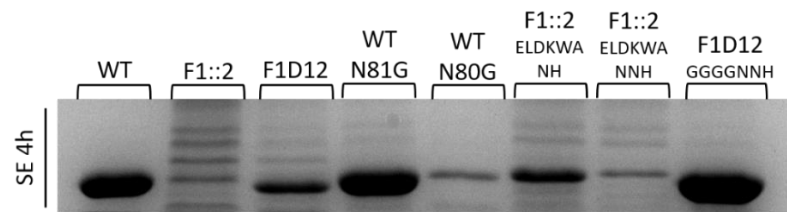


Figure 5.17: SDS-PAGE of surface F1 extracted from *E. coli* BW25113 to assess and improve export of loop 4 mutation.

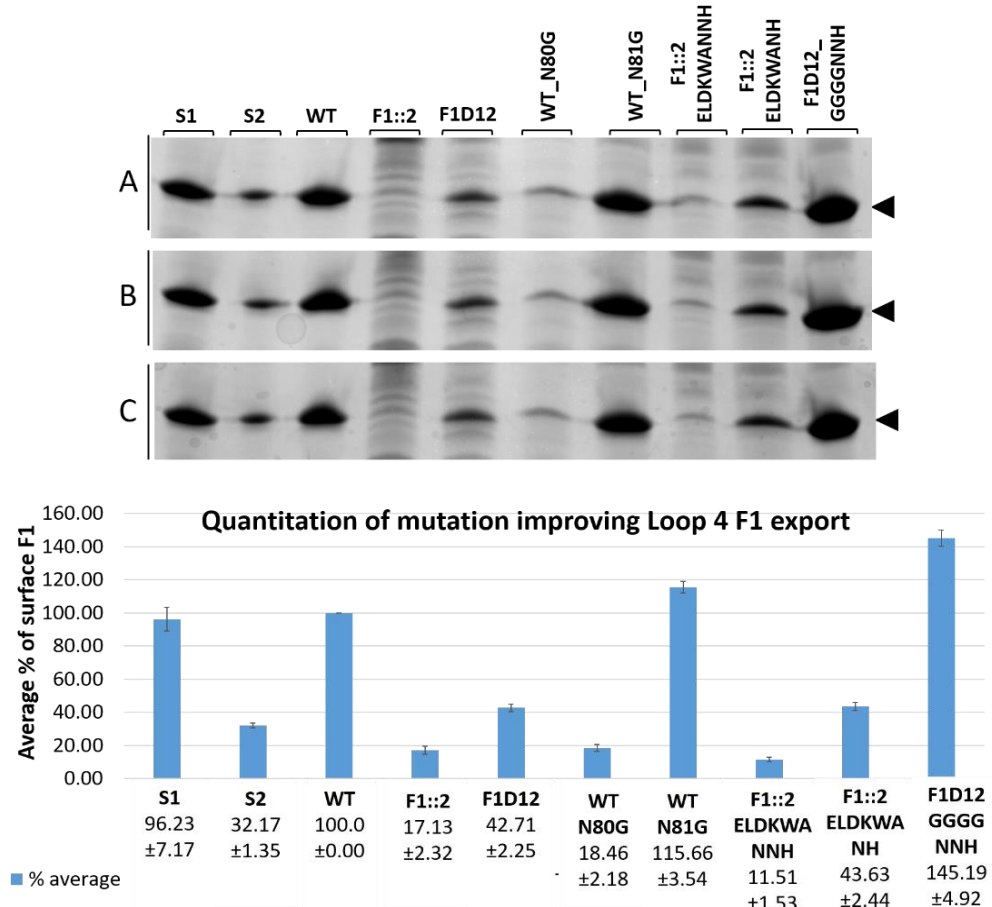


Figure 5.18: SDS-PAGE of extracted surface F1 from *E. coli*/pACYCF1 with point mutations in Caf1A.

Following 4 hour thermoinduction, F1 was extracted at 65°C for 1 h from 3 OD cells in 100µl PBS. Extracted samples (12µl) were mixed with 4µl SDS loading dye, heated at >95°C for 10 min to ensure complete denaturation of F1 fibre. Denatured sample (10µl) was loaded on SDS-PAGE. Following electrophoresis, protein was stained with Coomassie blue, imaged and bands of denatured Caf1 monomer (black arrowhead) quantitated. S1 and S2 indicate 2.59µg and 0.863µg, respectively, of the F1 standard applied. A, B and C are biological repeats, 3 separate transformants were processed for each construct.

5.5.3 Repairing clash and new interaction involving loop 5 of F1D13 and F1::1

Loop 5 gives the most promising level of modified F1 export for insertion of other epitopes. As shown in Figure 5.11, cells carrying F1::1 were estimated to export $47.9\% \pm 2.1$ of surface polymer compared to wildtype, the highest of any of the constructs carrying the charged epitope, ELDKWA. Modelling identified loss of an interaction between Ser93 (subunit) and Gln167 (Caf1A) (Figure 5.14). The significance of this was assessed by mutating Caf1AGln167 to a smaller residue Ala or Ser in wildtype F1. As shown Figure 5.19, mutation of this residue did not affect F1 export to the cell surface, suggesting that the interaction between subunit Ser93 and usher Gln167 alone is not critical. Hence, if need be Gln167 could be manipulated to improve F1 export.

Additionally, there was a clash between Caf1 Leu93 from the inserted epitope with Q169 in the Caf1A barrel. This was initially assessed by mutating Q169 to the smaller residues Ala and Gly with wild-type subunit. The cells carrying pACYCF1_Q169G or pACYCF1_Q169A both produce a similar level of F1 as wildtype usher, $112.10\% \pm 5.88$ and $100.42\% \pm 8.2$, respectively (Figure 5.19). This indicates that this residue is not essential and can be manipulated in an attempt to improve F1::1 production. The same mutations were prepared in pACYCF1::1. Both constructs showed dramatically improved production of F1::1; $86.54\% \pm 3.86$ from pACYCF1::1_Q169G and $78.71\% \pm 1.47$ from pACYCF1::1_Q169A compared to only $37.41\% \pm 1.26$ from pACYCF1::1 (Figure 5.11). Thus, in this case, a clash of residue was repaired by mutating Gln169 to a smaller residue. To confirm that the charged epitope insert was still in this construct, pACYCF1::1_Q169G was sent for whole

operon sequencing and both the sequence encoding the charged epitope with Caf1 subunit and the Q169G codon within *caf1A* were confirmed on this plasmid.

In the F1::1 usher translocon, a new interaction between Asp94 of inserted epitope and Ser182 of the usher β -barrel was modelled. To determine if this new interaction in the F1::1 translocon has any impact on the level of F1::1, Ser182 was mutated to Ala in pACYCF1 and also in pACYCF1::1. Cells carrying pACYCF1_S182A showed no impact on the level of surface F1, but cells carrying pACYCF1::1_S182A produced the same level of F1 as wildtype. This mean that the production of surface F1::1 increased from only $37.41\% \pm 1.26$ in cells carrying pACYCF1::1 to $110.68\% \pm 10.76$ in cells carrying pACYCF1::1_S182A. Time limitation, has not permitted whole operon sequencing of this construct. But based on the evidence provided, in contrast to Gln169 which appears to hinder export via a clash of residues, Ser182 appears to hinder export by interacting with a loop residue.

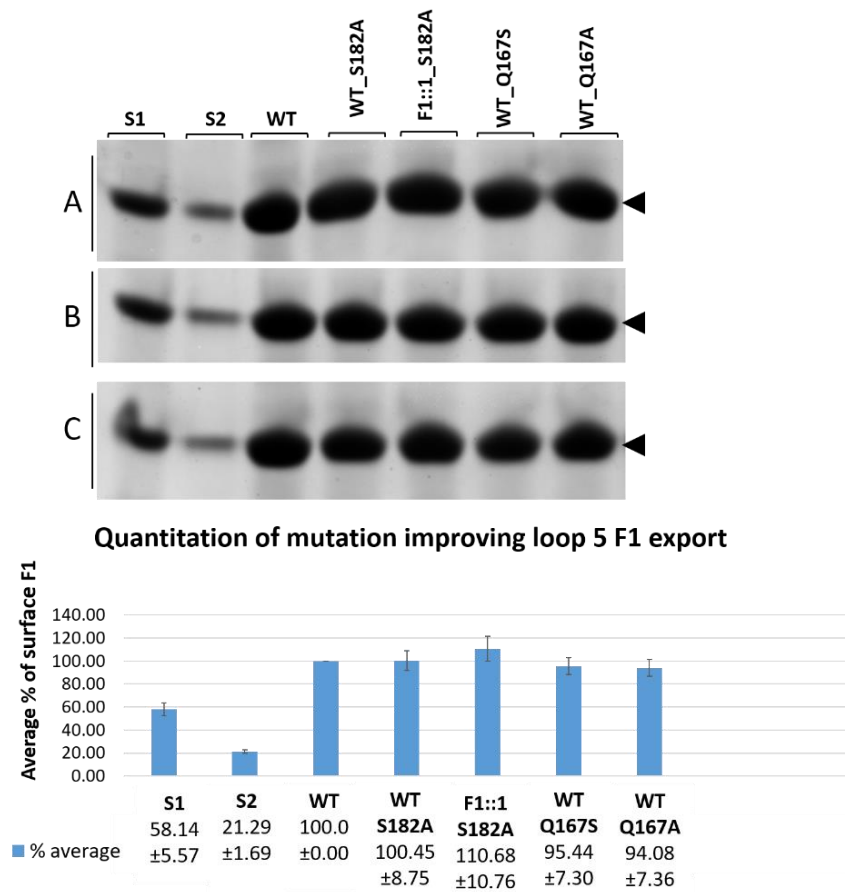
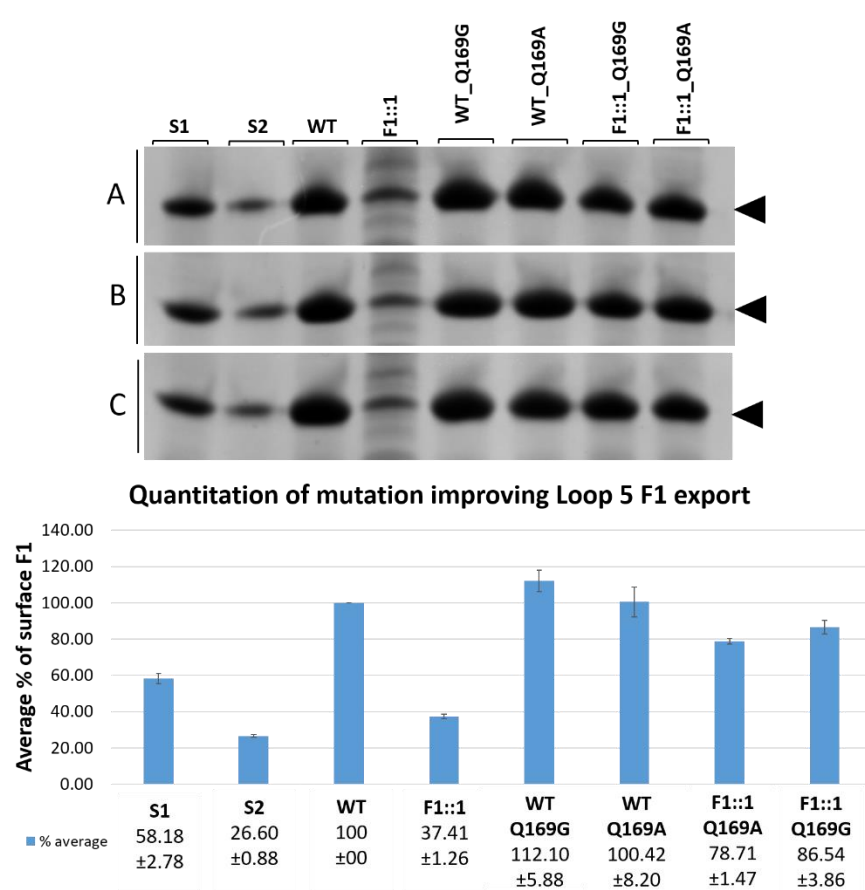


Figure 5.19: SDS-PAGE of extracted surface F1 from *E. coli*/pACYCF1 with point mutations in Caf1A.

Following 4 hour thermoinduction, F1 was extracted at 65°C for 1 h from 3 OD cells in 100µl PBS. Extracted samples (12µl) were mixed with 4µl SDS loading dye, heated at >95°C for 10 min to ensure complete denaturation of F1 fibre. Denatured sample (10µl) was loaded on SDS-PAGE. Following electrophoresis, protein was stained with Coomassie blue, imaged and bands of denatured Caf1 monomer (black arrowhead) quantitated. S1 and S2 indicate 2.59µg and 0.863µg respectively of the F1 standard applied. A, B and C are biological repeats, 3 separate transformants processed for each construct.

5.5.4 Repairing loss of interaction in loop 7 of pACYCF1D14 and pACYCF1::4 constructs

F1::4 is produced from pACYCF1::4 at very low levels, only $8.9\% \pm 1.15$ of the WT level of F1. Therefore, with F1::4 (loop 7), the focus was to remove the additional interaction that occurred between Asn80 and Asn81 with Tyr 287 and maintain an interaction between Asn80 and Tyr266. Asn81 was mutated to a smaller residue, Gly in pACYCF1::4. As shown in Figure 5.20, it can be seen that cells carrying pACYCF1::4_N81G abolished the production of surface F1 and no high molecular weight polymer was observed. Therefore, no further mutation was done to improve this construct as loop 5 had already been identified as the most promising loop to insert foreign epitope.

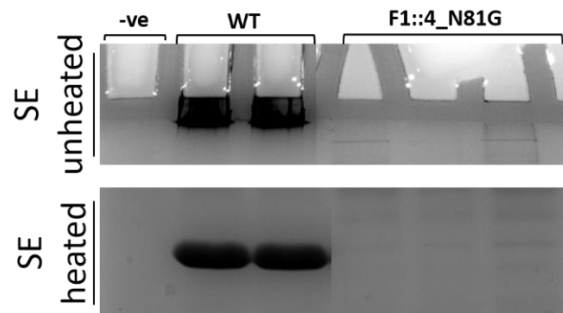


Figure 5.20: SDS-PAGE of surface F1 containing F1::4_N81G.

F1 extracted at 65°C for 1 hour and run on SDS-PAGE showing unheated sample with high molecular weight polymer in stacking gel. The heated sample was mixed with SDS loading dye and heated at >95°C for 15 minutes to denature all F1 fibre to Caf1 monomer. 3 OD cells after 4 hour thermoinduction were recovered, extracted and resuspend in 100µl PBS.

5.6 Histidine insertion in loop 1

In this study, loop 5 was also modified to insert six or four histidine residues, to develop a His tag within the surface assembled polymer. A His-tag within the fibre could have many applications, including detailed microscopy analysis of the fibre assembly. A construct was made by replacing the charged epitope residues from pACYCF1::1 with six His codons to create pACYCF1_DS6His. The Asp and Ser residues were also included to optimise loop flexibility for affinity labelling. InFusion HD cloning was used, by undergraduate students under my supervision, to make this and additional constructs.

The efficiency of assembly of F1::DSH₆ was tested by transforming pACYCF1_DS6His into *E. coli* Top10. As shown in Appendix 15 (A), F1::DSH₆ was produced and assembled into HMW polymer, but levels were substantially lower than that with wild-type F1. Production of his tagged F1 from cells carrying pACYCF1_DS6his was also evidently toxic after 4 hours thermoinduction. From the translocon model, this inserted epitope also clashed with Gln169. Therefore, Gln169 was mutated to Ala in this construct making pACYCF1_DS6his_Q169A. The level of F1 was tested by surface extraction and as shown in Figure 5.21 below the level of surface F1 did not improve after repairing the clash and the cells were still toxic after 4 hours induction at 37°C. The bulky His residues and number of inserted residues (extra 3 residues, from DSRDF to DSHHHHHH) was likely to be toxic. Thus, two histidine residues were deleted using Quickchange II XL kit making pACYCF1_DS4his to test the for cell toxicity and ability of binding to the Ni-NTA fluorescence probe.

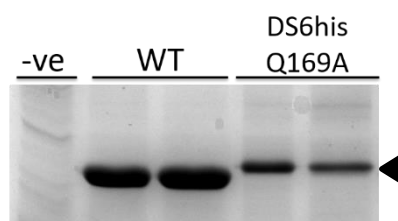


Figure 5.21: SDS-PAGE of surface F1::1DS6His.

E. coli Top10/ pACYCF1_DS6his_Q169A or pACYCF1 (WT) were thermoinduced for 4 h. F1 was extracted at 65°C for 1 h from 3 OD cells in 100µl PBS. Samples were mixed with SDS loading dye and heated at >95°C for 15 minutes to fully denature F1 fibre. Arrowhead, denatured Caf1M monomer.

A growth curve of pACYCF1_DS6his_Q169A and pACYCF1_DS4his constructs were tested for cell toxicity. The cells carrying these constructs were grown overnight at 26°C and then following 1 in10 subcultured were thermoinduced at 37°C. The OD₆₀₀ was monitored and whole cells sampled for F1 extraction, hourly.

Cells carrying pACYCF1_DS6his_Q169A grew more slowly and appeared toxic after 4 hours thermoinduction (Figure 5.22). The level of His tagged F1 from cells carrying pACYCF1_DS6his_Q169A, was very low initially - at 2, 3 and 4 hours. However, after 5 hours thermoinduction, the F1::DSHis level was almost the same as wildtype. This suggests that cells were recovering. For cells carrying pACYCF1_DS4his, the production of F1 was high compared to pACYCF1_DS6his_Q169A and this was supported by surface extraction (Appendix 15 (B)).

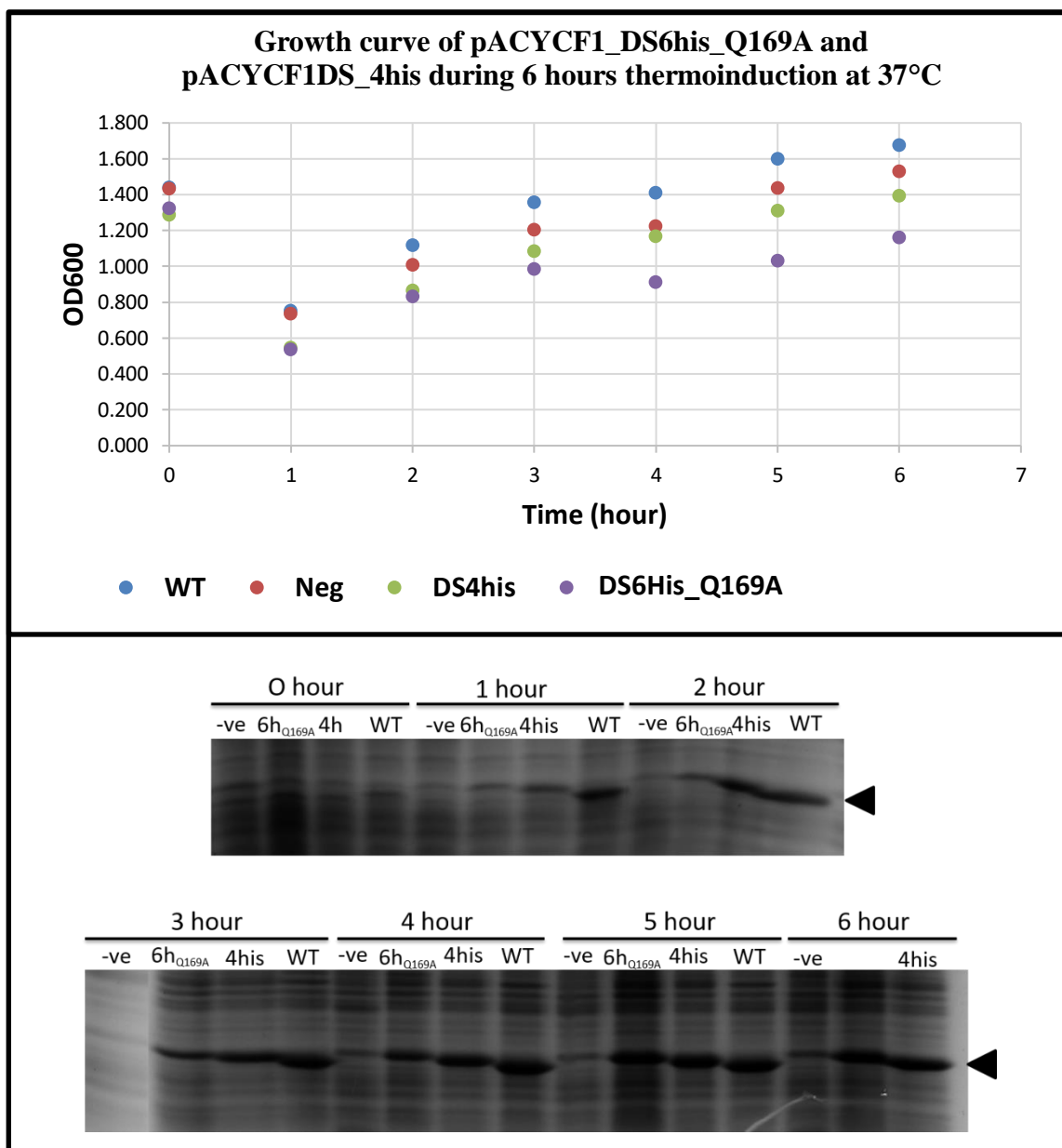


Figure 5.22: Growth curve of pACYCF1_DS6his_Q169A and pACYCF1_DS4his. (A) Graph showing growth curve of cells carrying control (pACYCDuet1), wildtype (pACYCF1), pACYCF1_DS6his_Q169A and pACYCF1_DS4his based on OD₆₀₀/ml taken every hour. (B) SDS-PAGE of F1 extracted from whole cells. 1 ml of cells of every hour thermoinduction were recovered and resuspended in PBS of 4OD/ml. *Note that spilled over of Caf1 monomer occurred in negative control lane.*

5.7 Conclusions

To test the ability of Caf1 to act as a scaffold for surface epitope, epitopes of the Caf1 subunit were predicted and loops 1, 4, 5 and 7 identified with strong B-cell epitope prediction. These loops were used to insert the charged epitope, ELDKWA and at the same time, the residues in this region were replaced with Gly to test the requirement of residues in each loop for efficient assembly through the translocon. The studies showed that all Gly loop replacements were exported and polymerised. However, F1D12 (loop 4) and F1D14 (loop 7) resulted in lower levels of modified surface F1 and F1D11 was less stable, already fully denatured at 75°C in SDS. Temperature stability tests showed that only F1 fibre from Gly replacement in loop 5 (F1D13) was fully stable, produced in high levels and behaved similarly to wildtype, and fully denatured after heating at 80-85°C in SDS. It is clear that insertion of the charged epitope, ELDKWA caused problems. The loop 5 insertion, expressed from pACYCF1::1 was the most permissive, accumulating 47.9% the level of wildtype F1. This construct produced high molecular weight polymer and represents the most promising site for insertion of short foreign epitopes.

In order to optimise the export of surface F1, translocon models were prepared for each modified F1. Each of the translocons was assessed to identify any residues that might interfere with polymerisation and export, with the aim of repairing problems to enhance export. One of the most striking outcomes of this was repair of the clash between Caf1A Gln169 and Leu of the ELDKWA sequence in F1::1 (loop5). By decreasing the size of Caf1AGln169, the level of surface F1::1 increased from only

37.41%±1.26, compared to wildtype F1, to 78.71%±1.47 and 86.54%±3.86 from cells carrying pACYCF1::1-Q169A and pACYCF1::1-Q169G, respectively.

A second important finding was the requirement of a correctly located Asn residue in loop 4. While the translocon model identified Asn81 as interacting with Tyr266 and mutation of Tyr266 led to a slight decrease in F1 assembly, subunit mutagenesis identified Asn80 as being the critical residue. In the F1 loop 4 mutations encoded on pACYCF1D12 and pACYCF1::2, Caf1Asn80 had been replaced with Gly or the charged epitope. F1D12 and F1::2 assembly was only 42 % and 6.9 %, respectively compared to wildtype F1. Unheated surface extracts of Caf1-F1D12 and Caf1-F1::2 showed a low level of high molecular weight polymer for Caf1-F1D12 and none in Caf1-F1::2. By replacing the Asn residues, modified F1 assembly, was restored to 145.19%±4.92 of wild-type F1, for pACYCF1D12. With F1::2 the size of the loop or orientation of residues was clearly important. Insertion of NH yielding the loop sequence ELDKWANH enhanced assembly from 17.13%±2.32% of wildtype to 43.63% ±2.44; whereas, insertion of NNH did not improve assembly. Location of an Asn residue (Asn80 in wildtype) is clearly essential for efficient surface assembly of Caf1.

Having demonstrated the ability to modify surface loops and achieve efficient export, an attempt was made to create an F1 polymer decorated with a 6His tag. As loop5 had been identified as the best loop to modify and loop4 clearly carried a required Asn, loop 5 was used for replacement with the polyhistidine tag. While 4 His residues were efficiently assembled, surface localisation of 6His residues were less successful and toxic to the cells.

CHAPTER 6 CHARACTERISATION OF

***E. COLI* SE11**

6.1 Introduction

E. coli SE11 is a commensal, non-pathogenic gram negative bacteria that was isolated from stools of a healthy adult. It has one chromosome (Genebank: AP009240) and six plasmids (Genebank: AP009241, AP009242, AP009243, AP009244, AP009245, AP009246) (Oshima et al., 2008). In chapter 4, an identity of 70% between an usher from *E. coli* SE11 and Caf1A was identified. This usher belongs to a fimbrial operon encoding a chaperone (ECSE_P2-0001), an usher (ECSE_P2-0002) and three putative subunits (ECSE_P2-0003, ECSE_P2-0004, ECSE_P2-0005) (see Figure 4.3). The operon is located on a 91.1 Kbp plasmid, pSE11-2 plasmid (AP009242). As discussed in Chapter 4, subunit 1 and subunit 2 from this operon have an identity of 30.4% and 31%, respectively, to Caf1. This was unexpected as Caf1 is considered unique to the pMT1/pFRA plasmids of *Y. pestis* and no close relatives had previously been identified on BLAST searches. This is particularly interesting for 3 reasons. Firstly, the backbone of pMT1 plasmid is already known to be related to other plasmids in the *Enterobacteriaceae*. Understanding the relationship between pMT1 and this SE11 plasmid may shed light on the acquisition of the *caf* operon by the early *Y. pestis* ancestor. Secondly, as both usher and subunits of the *Y. pestis* Caf locus and this *E. coli* SE11 CU system are more closely related than any other CU system, comparison of both sets of gene products should be advantageous in understanding CU assembly via the $\gamma 3$ subgroup. Finally, because of the sequence similarity between Caf1 and the putative subunits of these *E. coli* SE11 fibres, the possibility that there may be an immunological cross reaction between the two needs to be investigated. Any false positive in screening

for previous exposure to plague, could seriously interfere with assessment. This chapter outlines fundamental properties of proteins from the operon that are similar to Caf.

6.1.1 Comparison of *Y. pestis* pMT1/pFRA plasmid with commensal *E. coli* pSE11-2

Progressive MAUVE algorithm was used to compared plasmids between *Y. pestis* KIM5, *Y. pestis* CO92, *E. coli* SE11 and *S. enterica* serovar Typhi. *S. enterica* was included based on data in a publication in 2001. There is 57.6% and 56.2% similarity, respectively, between plasmid pFra (96.2 Kbp) from *Y. pestis* CO92 and pMT1 (100.9 Kbp) from *Y. pestis* KIM5 and pHCM2 (106.5 Kbp) from *S. enterica*. More than 90% of the similar regions shared 96% DNA identity (Prentice et al., 2001). To start, plasmid sequences were downloaded from the ftp site NCBI in genebank format for each plasmid. Based on the MAUVE results shown in Figure 6.1, pHCM2 does not shown any similarity with *caf* operon. Despite the low similarity of ECSE_P2 plasmid from *E. coli* SE11, there is a region (lime block in Figure 6.1) that is similar to *caf* operon. The white area in the yellow block of ECSE_P2 plasmid (91.1 Kbp) are the three subunits which are less similar and only have less than 35% identity with *cafI*. However, chaperone and usher shown to have a high similarity, as discussed above.

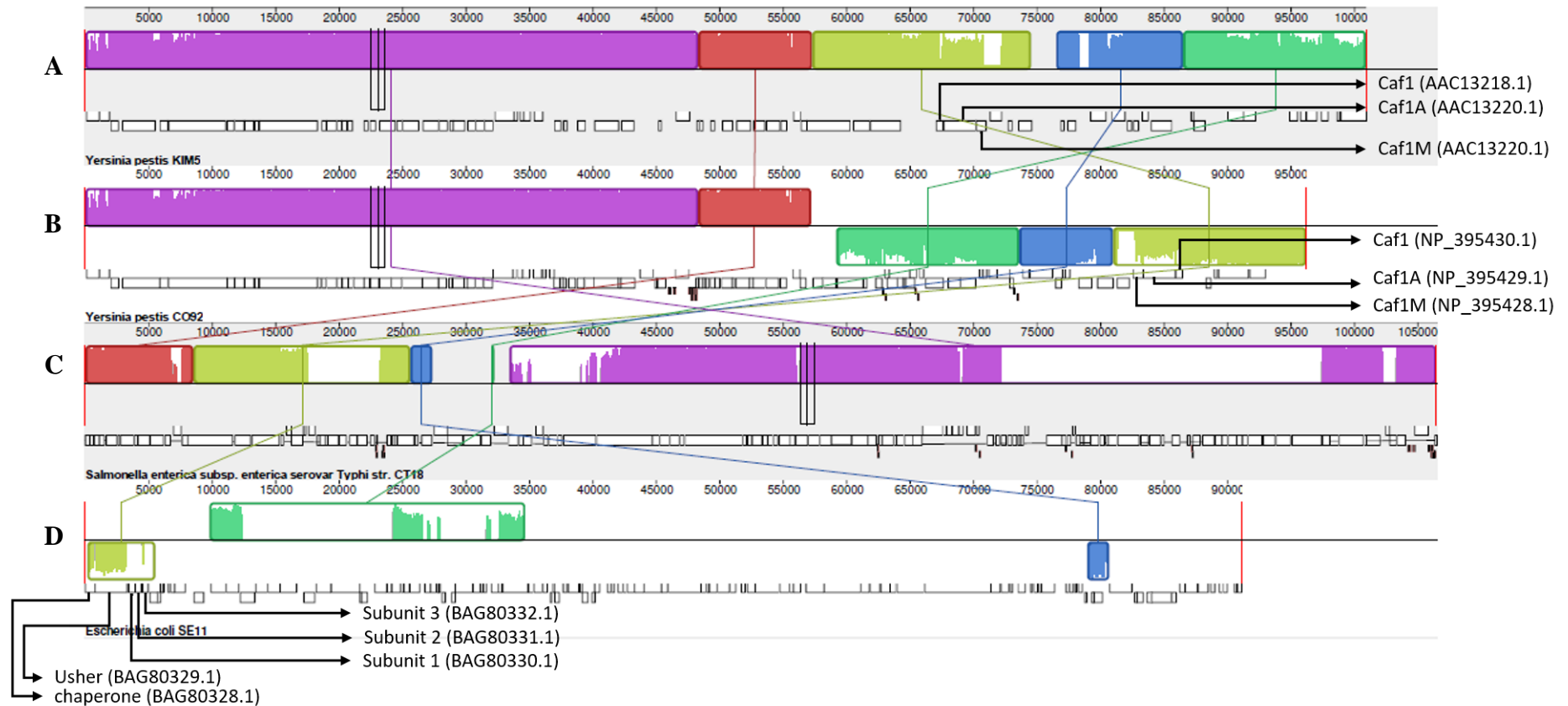


Figure 6.1: Comparison of plasmids.

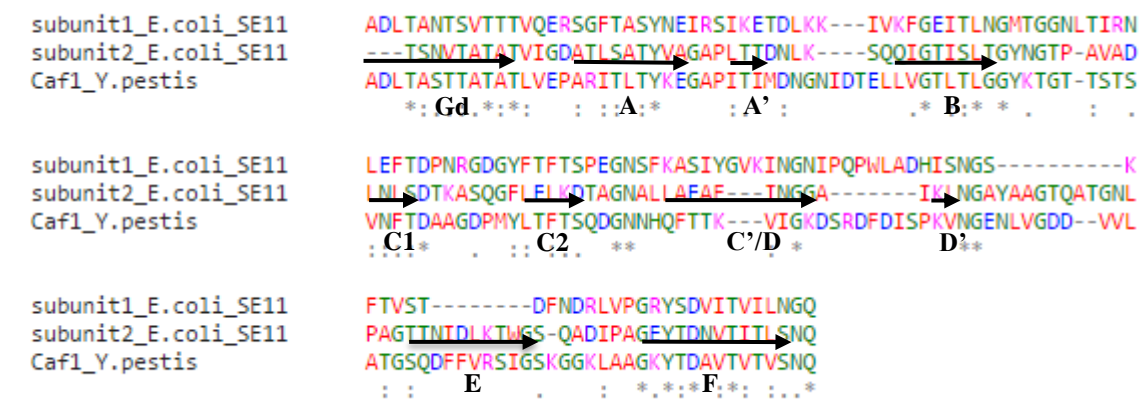
(A) *Y. pestis* KIM5 (pMT1, AF053947), (B) *Y. pestis* CO92 (pFra, NC_003134), (C) *S. enterica* (pHCM2, NC_003385) and (D) *E. coli* SE11 (ECSE_P2 plasmid, NC_011413). Colors depict conserved and highly related genomic regions and white areas identify unique or low identity regions. Blocks shifted below the centre line indicate regions that align in the reverse complement (inverse) orientation relative to reference *Y. pestis* KIM5.

6.1.2 Pairwise alignment with Caf1 and multiple sequence alignment of subunits

Analysis was performed on mature subunits. Therefore, SignalP programme was used to predict signal peptide cleavage. An unequivocal SP cleavage site was identified for both subunits. This was between Ala25 and Ala26 of subunit 1, leaving a mature protein of 134 amino acids and between Ala22 and Ala 23 of subunit 2 leaving a mature protein of 136 amino acids (Figure 6.3). The signal sequences were removed to compare mature subunits in pairwise alignments. The identity and similarity percentage of mature Caf1 (mCaf1) to mature *E. coli* SE11 subunit 1 is 30.4% and 45.6%, respectively; to subunit 2 it is 29.1% and 48.3%, respectively; and to subunit 3, 14.1% and 22.8% respectively.

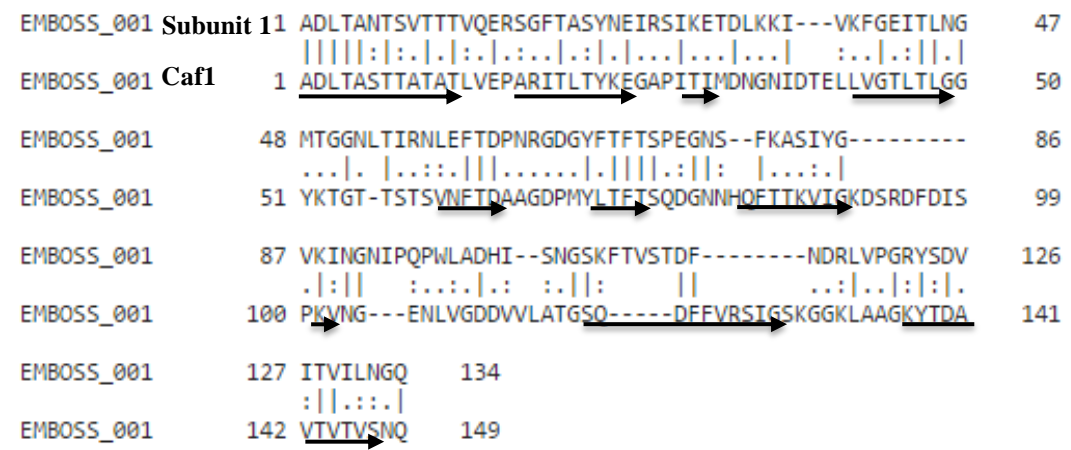
Multiple sequence alignment of mature Caf1, subunit 1 and subunit 2 was performed, omitting subunit 3 due to its short size (58 amino acids instead of 149 amino acids in subunit 1 and 2). This alignment (Figure 6.2 (A)) identified 11 conserved residues in β -strands and seven in the loops with high conservation in the Gd and F β -strands which are critical for polymerisation (MacIntyre, 2001; Zavialov et al., 2002; Yu et al., 2012).

A

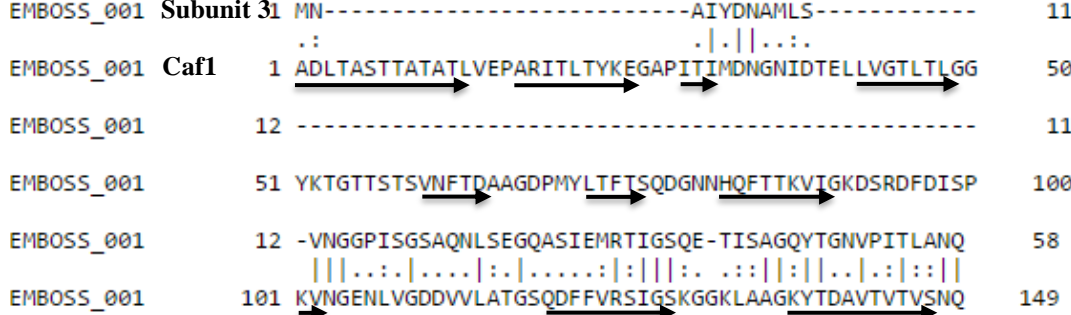


B

E. coli SE11 Subunit 1



E. coli SE11 Subunit 3



E. coli SE11 Subunit 2

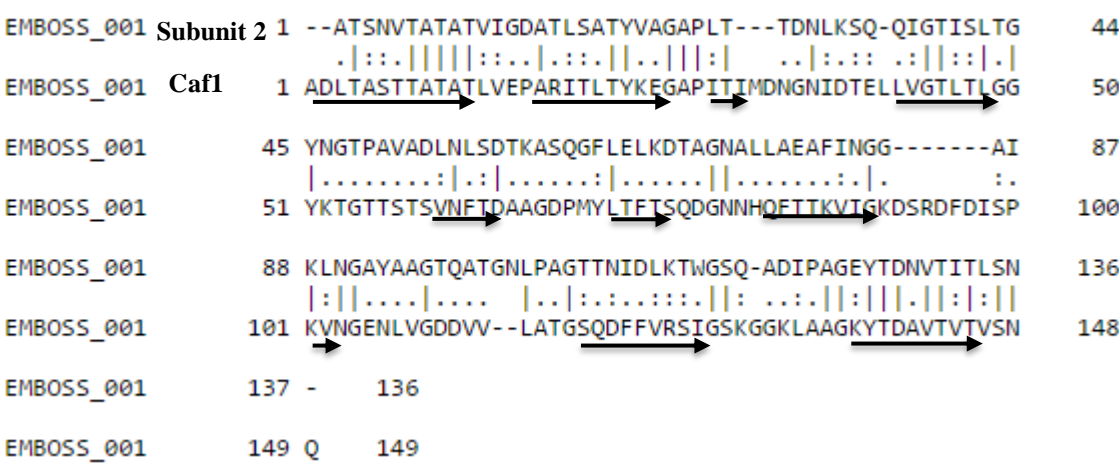


Figure 6.2: Multiple sequence alignment (ClustalW) and pairwise sequence alignment (EMBOSS) of *E. coli* SE11 subunits with Caf1. (A) Arrow indicates beta strands in Caf1. An asterisk (*) indicates conserved residues, colon (:) indicates conservation between residues with strongly similar properties, period (.) indicates conservation between residues with weakly similar properties. Mature Caf1 is aligned with precursor of all three subunits from *E. coli* SE11. (B) Pairwise sequence alignment. Top sequence is *E. coli* SE11 subunit and bottom sequence is Caf1, *Y. pestis*.

Interestingly, the important residue for stabilization of F1 polymer, Ala9 in Gd strand and surrounding residues are conserved between subunit 2 of *E. coli* SE11 and *Y. pestis* Caf1; whereas, the extreme N-terminus of subunit 1 is conserved. Notably, the P5 position (Gd strand Leu 13 in Caf1) is not present in either of the 2 SE11 subunits and is replaced with a conserved Val in both. This residue is important in initiating polymerisation (Yu et al., 2012). Both of these subunits were predicted to have capsule antigen Caf1 domain and hence would be predicted to form polymers of immunoglobulin-like modules via donor strand complementation.

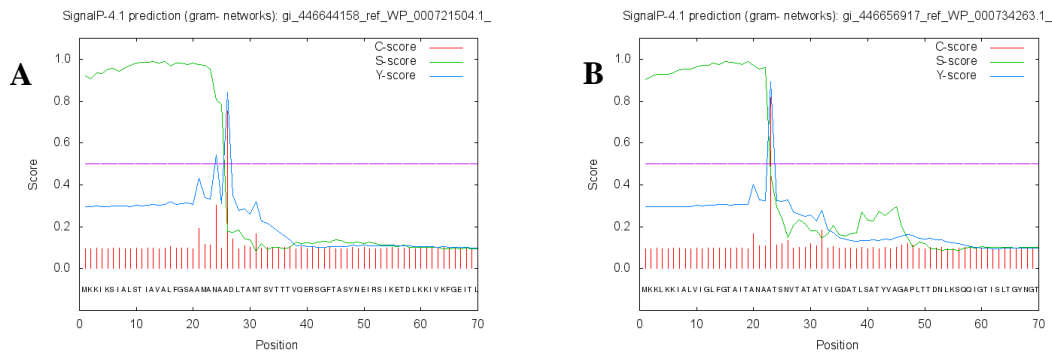


Figure 6.3: Prediction of signal peptide cleavage using SignalP.

(A) Subunit 1 (ECSE_P2-0003) from *E. coli* SE11 where the cleavage is between Ala25 and Ala26. (B) Subunit 2 (ECSE_P2-0004) where the cleavage is between Ala22 and Ala23. The C-score was used to distinguish signal peptide cleavage site, S-score to distinguish positions of signal peptides from the mature part of the proteins and from proteins without signal peptides and Y-score is combination of C-and S-score for better prediction.

6.1.3 Homology modelling of *E. coli* SE11 subunits and identification of conserved residues compared to Caf1

Following analysis of the *E. coli* SE11 subunit protein sequences, homology modeling was done for an in-depth comparison with the Caf1 subunit. I-TASSER was used for homology modeling using Caf1 (PDB:1Z9S:B) as template. The subunit models of *E. coli* SE11 and structure were then superposed using TM-align to compare the size and the shape. The model was used to locate conserved residues in the protein. As shown in figure 6.4, five of the 18 conserved residues, Gly50, Asn80, Gly90, Thr143 and Gln149, are surface located and facing outwards. Furthermore, Asn80 have been mutated in Caf1 and shown to be important for efficient export of F1 (Chapter 5).

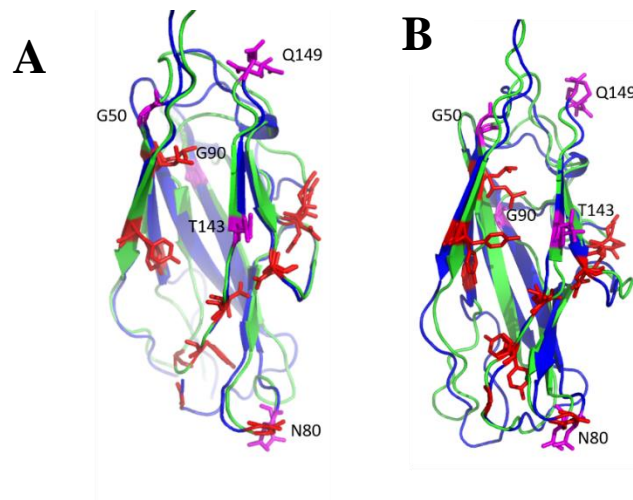


Figure 6.4: Modelled *E.coli* SE11 fibre subunits.

(A) Subunit 1 (ECSE_P2-0003), (B) Subunit 2 (ECSE_P2-0004). Both models are superimposed on structure of chaperone-bound Caf1 (PDB:1Z9S:B) Blue, Caf1; Green, *E. coli* SE11 Subunits; residues in red, conserved residues, facing inside; residues in pink are five conserved residues facing out and might be interacting with β -barrel Caf1A.

Table 6.1: Summary of conserved residues in Caf1 based on MSA of Caf1 and *E. coli* SE11 subunit 1 and 2.

Residue	Location
Thr4	Gd
Thr10	Gd
Thr12	Gd
Tyr23	A
Gly44	B
Leu48	B
Gly50	Loop 4
Asp64	Loop 5
Gly79	Loop 6
Asn80	Loop 6
Gly90	D
Asn103	Loop 8
Gly104	Loop 8
Gly136	F
Tyr138	F
Asp140	F
Thr143	F
Gln149	Loop 9

6.1.4 Prediction of B cell epitopes in *E. coli* SE11 fimbrial subunits

In section 5.2, four epitopes in Caf1 subunits were predicted using DiscoTope. Discotope was also used to predict B cell epitopes in *E. coli* SE11 subunits 1 and 2 and MSA to compare location of predicted epitopes in Caf1 and each *E. coli* SE11 subunit. Based on Figure 6.5, subunit 1 has 8 regions predicted as epitope in the mature protein. Three of these, in loops 1, 4 and 5 correspond to locations of predicted epitope in Caf1 and loop 5, possessed the conserved residues, Gly79 and Asn80. In contrast, subunit 2 shared no location of predicted epitope with Caf1. In fact, subunit 2 has only a single predicted epitope corresponding to loop 1 and a weak prediction for the Gd strand. This would suggest

that, if there is any cross-reaction between F1 and polymers from this *E. coli* SE11 CU system, then there is a greater chance of subunit 2 cross reacting with antibody against F1.

6.2 Construction and characterisation of recombinant *E. coli* SE11 CU operon

6.2.1 Cloning *E. coli* SE11 *caf*-like operon for recombinant fibre assembly and tagged chaperone-subunit complex

To experimentally test functionality and any cross-reaction between CU fibres of the *E. coli* SE11 locus and *Y. pestis* F1, the complete pSE11-2 CU operon was cloned into pACYCDuet1 to test for surface assembly of an SE11 fibre. Chaperone and subunits, without usher, were also cloned into pTRC99 to test for recovery of periplasmic polymer. The constructs pACYCD1-MA-S123 and pTRC99-MHis-S123 (Figure 6.6) were created using InFusion Cloning (Takara). For construction of pACYCD1-MA-S123, pACYCDuet1 was linearised by restriction digestion, NcoI (NEB). The complete *E. coli* pSE11-2 fimbriae sequence from chaperone to subunit 3 (4997 bp) was amplified using CloneAMPTM HiFi Premix and primers . The Infusion Cloning kit (Takara) was used to InFuse the insert into the linearised vector, which was then transformed into *E. coli* Stellar ultra competent cells. Plasmid DNA was analysed for the correct size of 8.9 Kbp for pACYCD1-MA-S123 and. The entire insert and fusion joints confirmed by sequencing.

pTRC99-MHis-S123 was created by linearising pTRC99 by restriction digestion with NcoI (NEB). Two inserts were used to make this construct; fragment one contained DNA encoding the chaperone with a C-terminal 6Histag and fragment two contained DNA encoding, subunits 1, -2 and -3 (Figure 6.6). Both were amplified from *E. coli* SE11 genomic DNA using CloneAMPTM HiFi Premix and SE11For1 and SE11Rev1 primers. The amplified inserts and linearised vector were InFused and transformed into *E. coli* Stellar ultra competent cells. Individual transformants were picked, plasmid (6.7 Kbp) extracted

and the correct insert confirmed by sequencing. The CU system encoded in pACYCD1-MA-S123 is under control of the T7 promoter and the chaperone and subunits in pTRC99-Mhis-S123 are under control of the *pTRC* promoter. Both require induction with IPTG for expression.

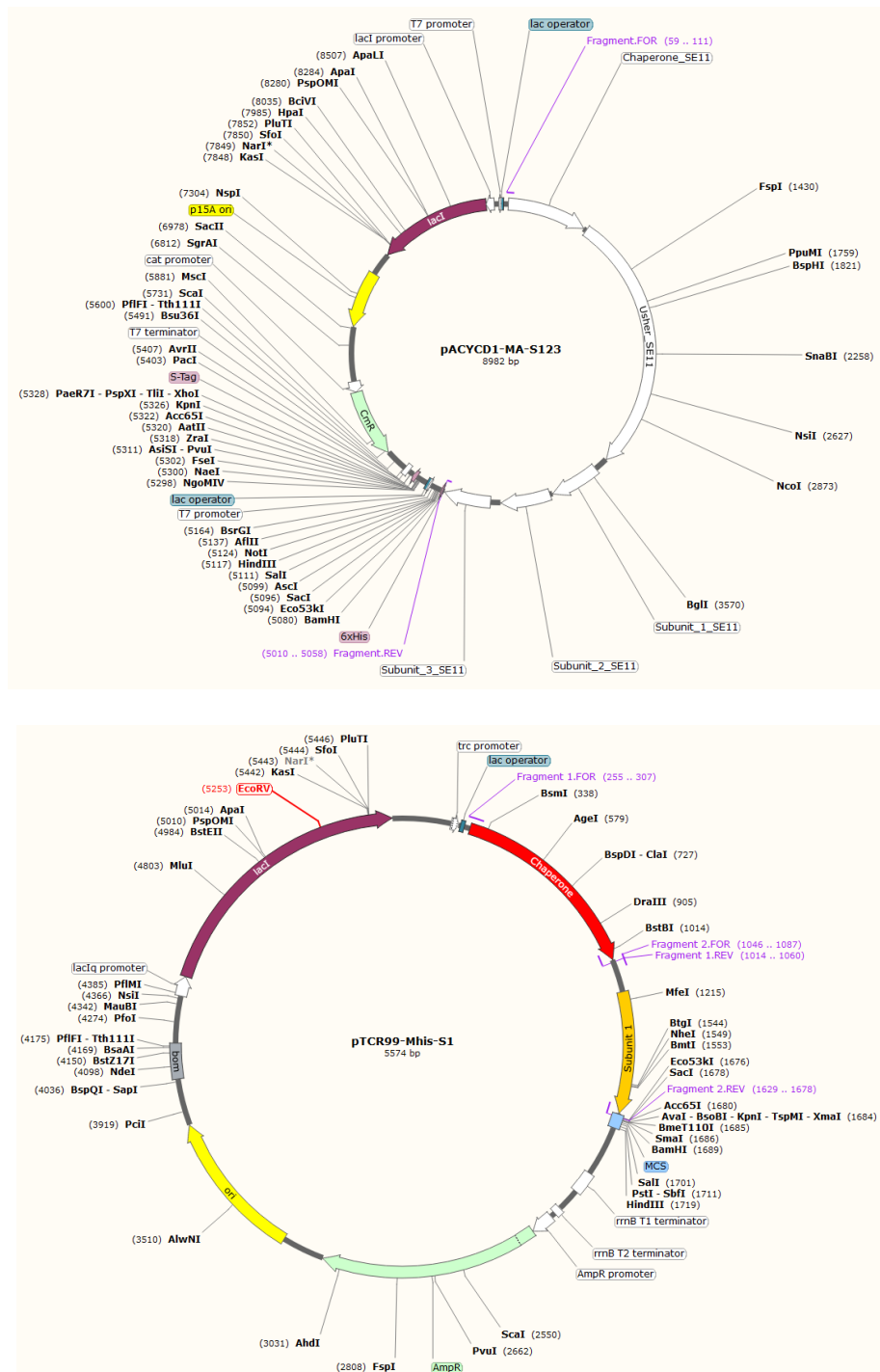


Figure 6.6: pACYCD1-MA-S123 and pTRC99-MHis-S123 constructs. pACYCD1-MA-S123 was used for recombinant surface expression of *E. coli* SE11 fimbriae, expressed from T7 promoter using plasmid SD sequence upstream of chaperone gene. pTRC99-MHis-S123 was used to trap the fimbriae in the periplasmic. The map was generated using SnapGene 3.2.1 tool.

6.2.2 Lack of toxicity of cloned CU operon and identification of expressed protein

To test for toxicity of expression of the recombinant CU system, pACYCD1-MA-S123 was transformed into *E. coli* BL21 (DE3) and growth monitored over 6 h induction. The cells were grown overnight at 26°C in LB with 0.6% glucose and 34 ug/ml CAM. The next day, the culture was washed and subcultured 1/10 into LB with 10 ug/ml CAM and grown at 37°C for 1 hour. The cells were then induced with 0.5 mM IPTG. The culture was sampled every two hours; the OD₆₀₀ was read and whole cells (1 ml) recovered and resuspended in PBS to a concentration of 4 OD cells/ml. Denaturated cells were then analysed by SDS-PAGE. Based on Figure 6.7, it is clear that cells carrying pACYCD1-MA-S123 grew at a similar rate to pACYCF1 and in fact better than the control, pACYCduet1. Expression of recombinant protein was confirmed by SDS-PAGE. In samples of cells carrying pACYCD1-MA-S123 recombinant protein was clearly visible even after 2 hours induction and the level increased at 4 and 6 hours induction. Two bands migrating slightly faster than Caf1 were observed. These were absent from the control and could be assigned as subunits from *E. coli* SE11. These results indicate that expression of the recombinant SE11 CU operon is not toxic to the *E. coli* cells, under the conditions tested. The amount of recombinant subunit was similar at 4 hours and 6 hours induction thus, 4 h induction with 0.5 mM IPTG at 37°C was used for further studies.

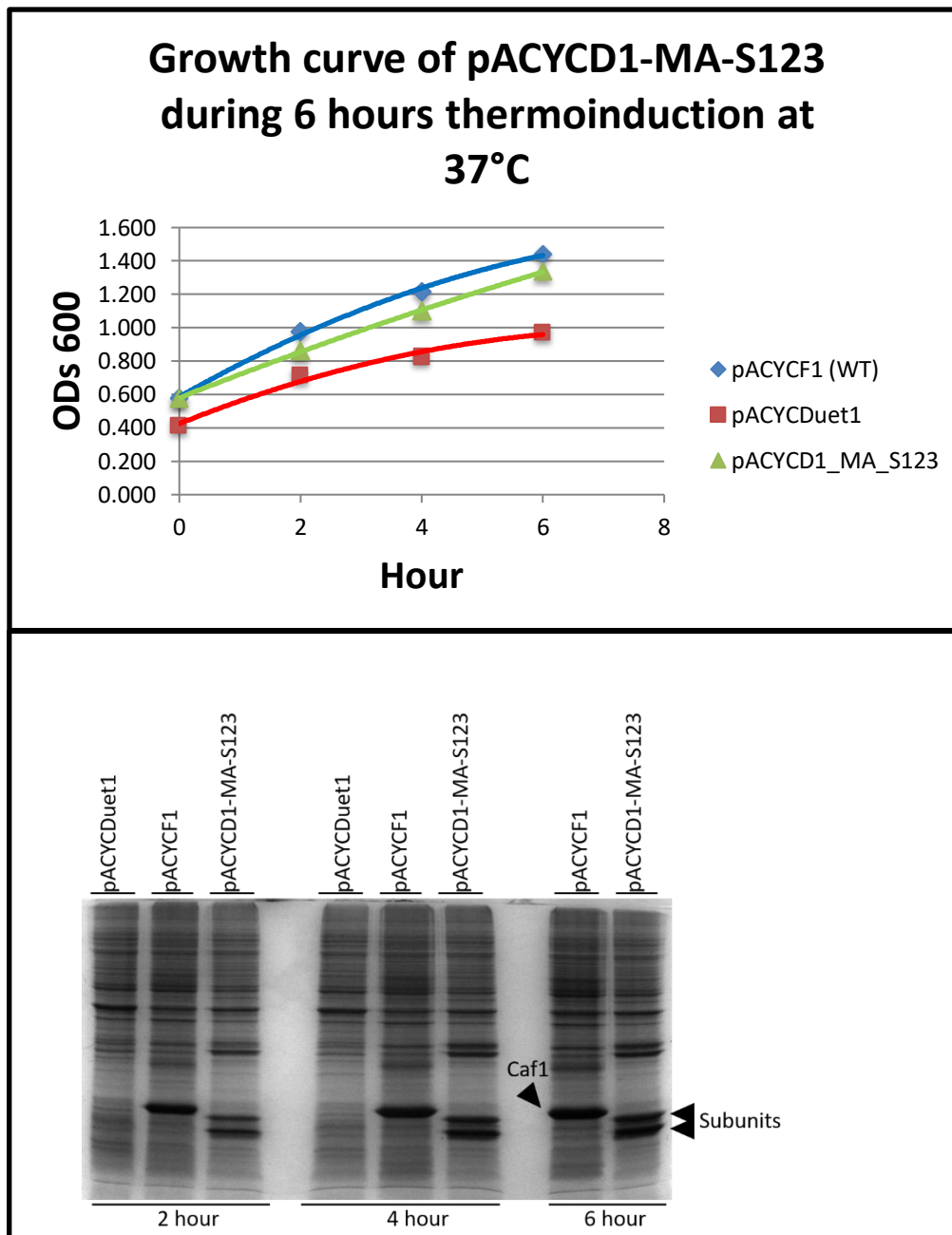


Figure 6.7: Growth curve of recombinant *E. coli* BL21 (DE3)/pACYCD1-MA-S123 (A) Growth curve of *E. coli* carrying pACYCDuet1 (empty vector), pACYCF1 (recombinant *Y. pestis* *caf* locus) and pACYCD1-MA-S123 (recombinant CU system from *E. coli* SE11). OD₆₀₀ were taken from each culture every 2 hours, after IPTG induction (0 hour, time) (B) SDS-PAGE of whole cells. Cell samples (1 ml) were recovered, every 2 hours after induction and resuspended in PBS, at a concentration of 4 OD cells/ml PBS and analysed by SDS-PAGE. Arrow indicates subunits from *E. coli* SE11 or Caf1. Note that at 6 hours, pACYCDuet1 sample was missed.

6.2.3 Comparison of recombinant production of periplasmic chaperone-subunit complexes and surface assembled subunit from the *E. coli* SE11 fimbriae operon

Expression of the recombinant *E. coli* SE11 fimbriae operon from pACYCD1-MA-S123 was not toxic to *E. coli* BL21 and subunits were produced in high level. Therefore, a comparison of the level of subunit accumulating on the surface from *E. coli* BL21 (DE3) /pACYCD1-MA-S123 and in the periplasm from *E. coli* Top10/pTRC99-Mhis-S123 was compared. His-tagged periplasmic subunit would be simpler to purify if expressed at a reasonable level. Both cultures were grown overnight at 26°C in LB with 0.6% glucose with appropriate antibiotics. The next day, the cultures were washed and subcultured 1/10 into LB and appropriate antibiotics. After an hour (OD₆₀₀ ~0.5), the cells were induced with 0.5 mM IPTG for 4 hours followed by extraction of surface polymer at 65°C/1 hour or isolation of the periplasmic fraction by osmotic shock (see section 2.8.2 and 2.8.3). As can be seen on the SDS-PAGE (Figure 6.8), two bands at approximately 13 kDa and 14 kDa were observed in both surface (pACYCD1-MA-S123) and periplasmic (pTRC99-Mhis-S123) extracts of the respective denatured samples. At the same time, surface extracts were prepared from *E. coli* SE11, using the same growth and extraction conditions. There was no striking evidence of expression of these proteins, a faint band was observed at the size of 13.5 kDa and several other higher molecular weight bands. At present, there is no evidence that these are related to the plasmid encoded CU system of interest. It is also quite likely that expression of this SE11 system is tightly control in the host bacterium.

In this experiment, pFM1 was used as control for periplasmic extraction as the high level of chaperone traps Caf1:Caf1M complexes resulting in accumulation of Caf1 polymer

in the periplasm. In the periplasmic of cells carrying pTRC99-MHis-S123, the level of chaperone appeared to be low and the level of trapped subunits lower than surface subunits. One conclusion here might be that the *E. coli* SE11 subunits behave in the same way as Caf1 and are degraded when not bound to the chaperone.

Unheated samples of cells carrying pACYCD1-MA-S123 were observed to have high molecular weight polymer where the polymer can be observed in the stacking gel and oligomers in the separating gel. This suggests that this polymer was stable even when the cells were heat shocked at 65°C to extract the surface protein. With periplasmic extracts, following expression from pTRC99-MHis-S123, the same HMW polymer was not noted, but the subunit bands disappeared, implying that they had also polymerised. Immunoblotting would be required to confirm this. Notably, the bands for both denatured subunits disappeared in the absence of heating, suggesting that both subunits form HMW polymer.

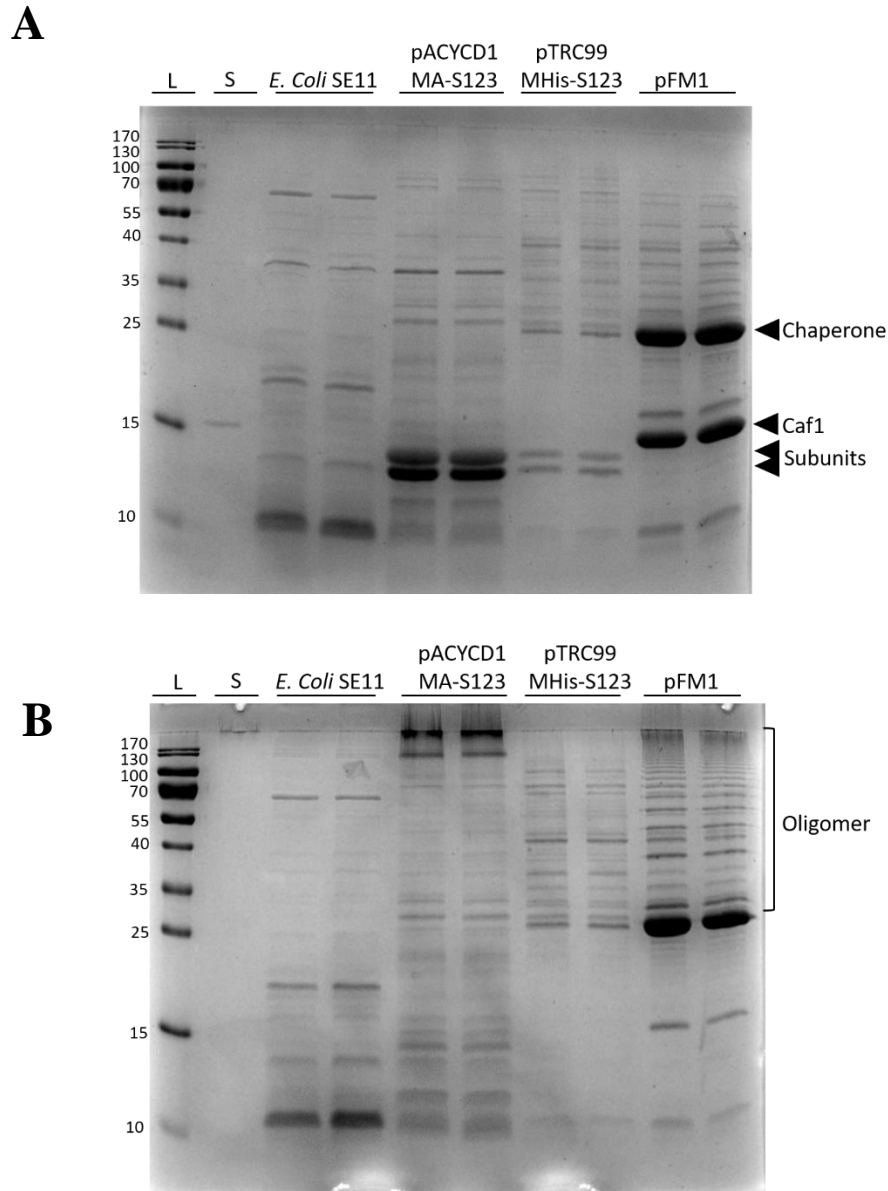


Figure 6.8: SDS-PAGE of surface and periplasmic subunits/polymer of *E. coli* SE11 fimbriae operon.

(A) Heated periplasmic and surface protein at 95°C for 15 minutes (B) F1/ *E. coli* SE11 polymer (stack gel) and Caf1/*E. coli* SE11 subunit oligomers (separating gel) in unheated samples. 3 OD cells after 4 hour induction were recovered and resuspend in 100µl PBS.

6.2.4 Heat stability of recombinant surface assembled polymer

To assess the stability of *E. coli* SE11 polymer, cells carrying pACYCD1-MA-S123 were heat shocked at 55°C to extract surface polymer and subjected to heat denaturation at 65°C, 75°C, 85°C and 95°C in SDS-PAGE loading dye as shown in Figure 6.9. The SE11 polymer was stable at 55°C where a high molecular weight polymer can be observed in the stacking gel. The polymer appears to denature at a temperature between 55 and 65°C. Both subunits are started to denature at 65°C with no HMW polymer in the stacking gel. Oligomers can be observed when the surface protein heated at 65°C, 75°C and 85°C and fully denatured at 95°C. This suggests that the polymer was very stable where it can only be fully denatured at 95°C in SDS. Although there is consistently less of the upper subunits, both subunits appear to behave in the same way.

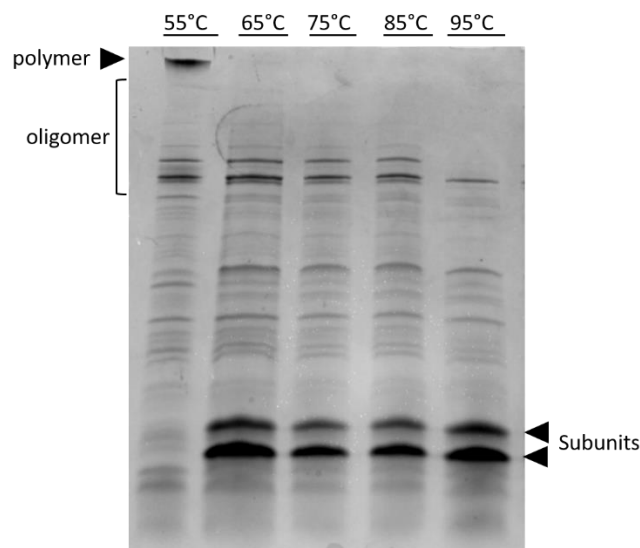


Figure 6.9: Temperature stability of surface extracted cells carrying pACYCD1-MA-S123. 4.5 OD cells after 4 hour thermoinduction were recovered and resuspended in 100µl PBS. Extracted surface protein (12 µl) was mixed with 4X SDS-loading buffer (4 µl) and heated for 10 minutes at 65°C, 75°C, 85°C, 95°C in an Eppendorf Master Cycle PCR machine. 10µl was loaded unto the well for all samples.

6.3 Mass spectrometry of *E. coli* SE11 subunit

Following SDS_PAGE, two protein bands were resolved at approximately 14 kDa and 15 kDa following expression from both pACYCD1_MA_S123 and pTRC99-Mhis-S123. To confirm identity of both bands, mass spectrometry following tryptic digests of both bands was performed. A large batch of *E. coli* BL21 (DE3) cells carrying pACYCD1_MA_S123 was induced for 4 hours and surface fimbriae heat-extracted at 65°C for one hour. The surface extracted samples were applied to a 16% acrylamide gel for SDS-PAGE, stained with Coomassie blue and destained. The bands migrating at approximately 14 kDa and 15 kDa (Figure 6.10) plus a control band of 70 kDa from the ladder and a background acrylamide gel control were excised, and processed for trypsin digestion and mass-spectrometry as described (Section 2.9.6). Identified peptides are listed in Table 6.3. The peptide profiles confirmed that the two bands were indeed the two different subunits (and not a breakdown product of one subunit). Based on the peptides identified by mass spectrometry, the upper band of 15 kDa protein was identified as subunit 2 and the lower band of 14 kDa protein was identified as subunit 1. It is obvious from all analysis thus far that there were only 2 major bands expressed from this locus. Subunit 3 would be much smaller and not big enough for a complete pilin subunit. From this analysis also, it can be confirmed that subunit 3 was not one of these bands. If expressed at all, it is not a major component of the fibre.

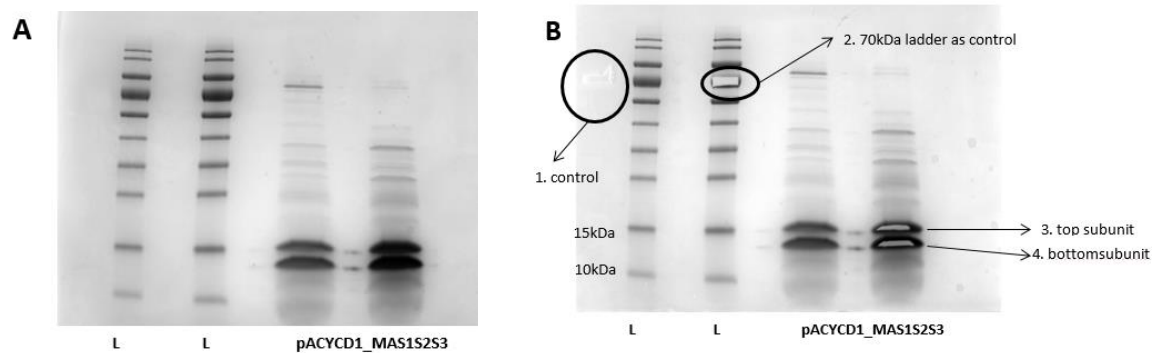


Figure 6.10: SDS-PAGE of surface extracted *E. coli* SE11 fimbriae. (A) SDS-PAGE before excising the gel. (B) SDS-PAGE after excising the gels. L indicates ladder.

Table 6.2: Peptides identified by mass spectrometry

A

Peptide	Subunit peptide found
RNLEFTDPNRRG	Subunit 1
KFTVSTDFNDRL	Subunit 1
RSGFTASYNEIR	Subunit 1
RYSDVITVILNGQ	Subunit 1
KFTVSTDFNDRLVPGR	Subunit 1
RGDGYFTFTSPEGNSFK	Subunit 1
KFGEITLNGMTGGNLTIR	Subunit 1
KINGNIPQPWLADHISNGSKF	Subunit 1
RNLEFTDPNRRGDGYFTFTSPEGNSFKA	Subunit 1

B

Peptide	Subunit peptide found
KDTAGNALLAEAFINGGAIKL	Subunit 2
KLNGAYAAGTQATGNLPAGTTNIDLKT	Subunit 2

6.4 Purification of recombinant subunits from *E. coli* SE11

6.4.1 Comparison of His tagged periplasmic chaperone complexes

Purification of *E. coli* SE11 subunits can be used for antibody production to test for any cross-reaction between these SE11 CU fibres and *Y. pestis* F1 polymer. Antibody will also be very useful to monitor expression from *E. coli* SE11. The C-terminal 6Histidine-tagged construct (pTRC99-MHis-M123) was created to trap *E. coli* SE11 subunits/polymer within the periplasm and facilitate purification. However, to optimise binding of the chaperone:subunit/s complexes to the Ni-NTA, a C-terminal 8Histidine-tagged *E. coli* SE11 chaperone was created using QuikChange II XL Site-Directed Mutagenesis by inserting 2 histidine to pTRC99-MHis-M123, making the construct pTRC99-M8His-S123. At the same time, to produce antibody against subunit 1, a construct of pTRC99-M6his-S1 was created by deleting subunit 2 and 3. Similarly, a construct of pTRC99-M6his-S23 was attempted by deleting subunit 1, for production of antibody against subunit 2. However, creation of this latter construct was unsuccessful. Therefore, pTRC99-M8his-S123 and pTRC99-M6his-S1 were processed for periplasmic extraction to compare the production of *E. coli* SE11 subunits/polymer. Figure 6.11 shows that cells carrying the pTRC99-Mhis-S1 construct produced only subunit 1 as expected and that cells carrying pTRC99-M8his-S123 produced both subunit 1 and subunit 2. Both constructs produced low levels but visible subunit. Therefore, both constructs were used for purification with an Ni-NTA column.

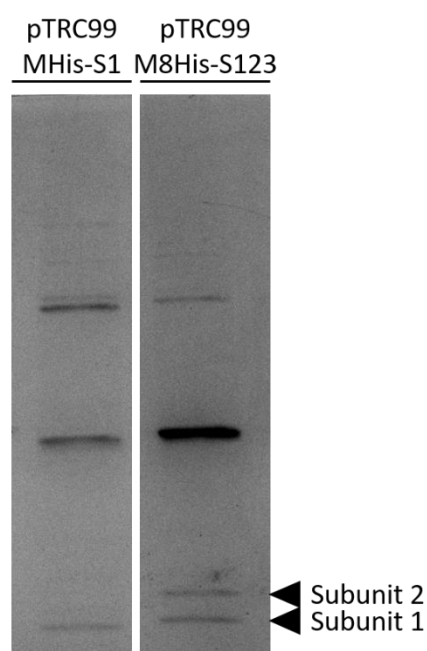


Figure 6.11: SDS-PAGE of periplasmic subunits/polymer from pTRC99-Mhis-S1 and pTRC99-M8His-S123.

After 4 hour induction with IPTG at 37°C, periplasm was prepared from 3 OD cells in a final volume of 100 ul 10 mM MgCl₂.

6.4.2 Purification of *E. coli* SE11 chaperone:subunit/s complexes by Ni²⁺ NTA

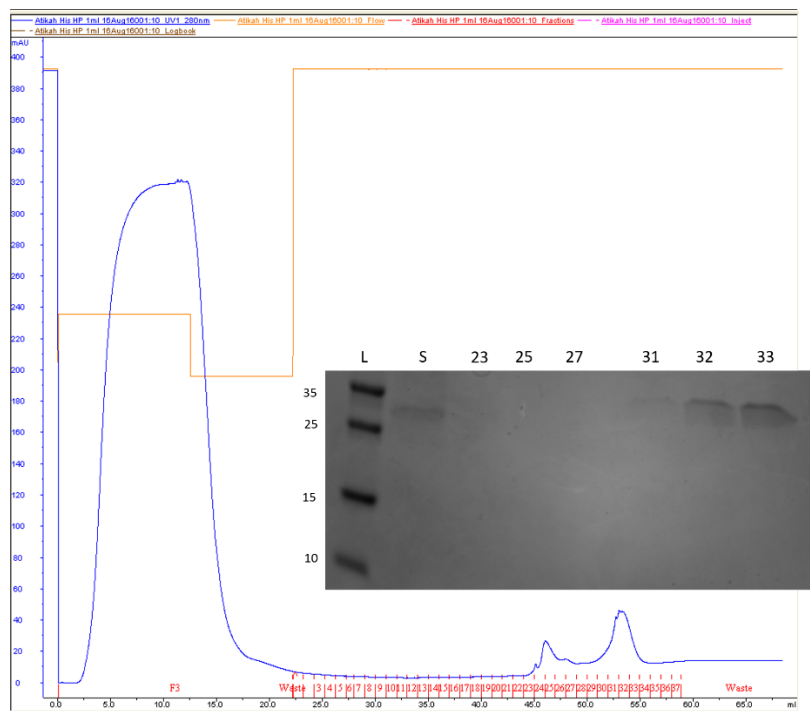
The large batch of periplasmic fraction, 9 ml from 400 ml induced cultures of *E. coli* BW25113/pTRC99-M6His-S1/DH5α and *E. coli* BW25113/pTRC99-M8His-S123/ DH5α, were applied to 1ml His-Trap columns and eluted over 30 fractions of 1 ml. Peak fractions were analysed by SDS-PAGE. Chaperones were observed at the size of approximately 30 kDa at fraction 31, 32, 33 and 21, 22, 23, 24 in pTRC99-M8His-S123/DH5α and pTRC99-M6His-S1/DH5α, respectively (Figure 6.12). This is in agreement with tighter binding of the 8 His tagged chaperone. However, no subunits were visible in stained gels from either purification. It could be that in subunit loaded chaperone the His tag is no longer accessible to bind the column. Due to this and the low level of subunits accumulating in periplasmic

fractions from these constructs, it was decided not to pursue this approach any further and to focus on surface assembled protein.

6.4.3 Purification of recombinant surface assembled *E. coli* SE11 subunits by size exclusion chromatography

Since the purification of *E. coli* SE11 chaperone:subunits/s complexes was unsuccessful, pACYCD1-MA-S123/BL21 (DE3) was used for purification. Heat extraction provides an excellent first step purification. This was further purified by treatment with TritonX 114 to remove LPS and lipids followed by size exclusion chromatography to separate HMW polymer from lower protein contaminants. The surface extracted sample (refer section 2.9.8 for procedure) was applied to a 120 ml HiPrep Sepharyl HR S-300 column in PBS and 90 fractions of 1.5 ml each were collected at 0.5 ml/min. Peak fractions were analysed by SDS-PAGE. Subunits 1 and 2 were eluted together in fraction 18 and 19. No other proteins were visible in these fractions or in any other fraction. Due to the low level of contaminating proteins, samples will be reanalysed following concentration and are being saved at -20C to raise antibody. As these subunits have an identity of more than 30% to the unique F1 antigen, it is important to exclude any immunological cross-reaction between the *E. coli* SE11 subunits and *. pestis* F1. Unfortunately, due to the time constraint, antibody production and cross reaction experiments cannot be done within this study. The antibody will also be useful to further characterise this commensal CU system.

A



B

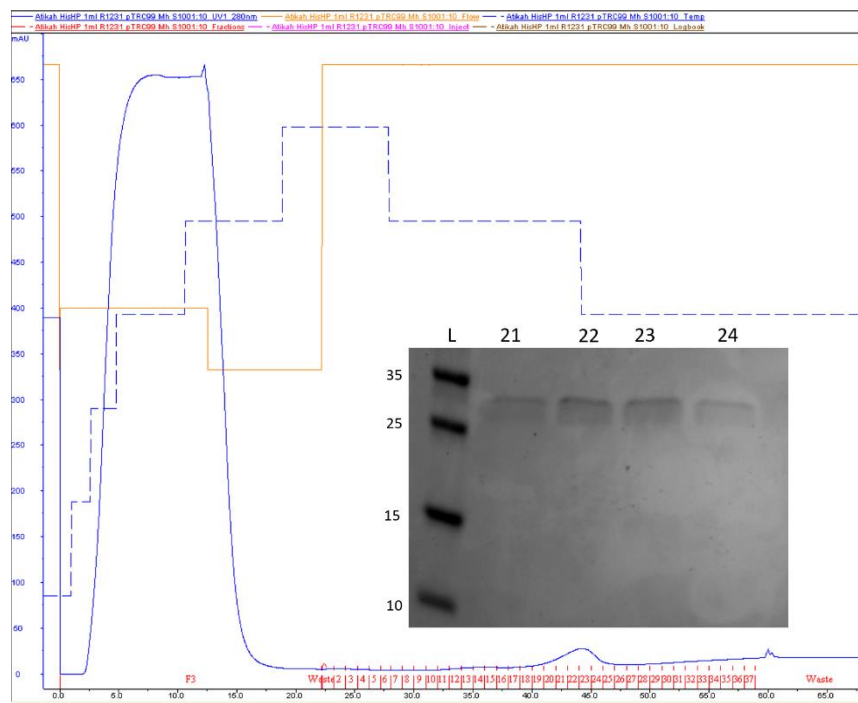


Figure 6.12: NTA purification of recombinant *E. coli* SE11 chaperone:subunits complexes.

(A) Chromatogram and SDS-PAGE of pTRC99-M8his-S123. (B) Chromatogram and SDS-PAGE of pTRC99-M6his-S1. Peak fractions were boiled for 15 mins at >95°C ins SDS-PAGE loading dye prior to SDS-PAGE. Number corresponds to eluted fraction number of 1ml each fractions. S indicates sample before purification and L indicates protein ladder.

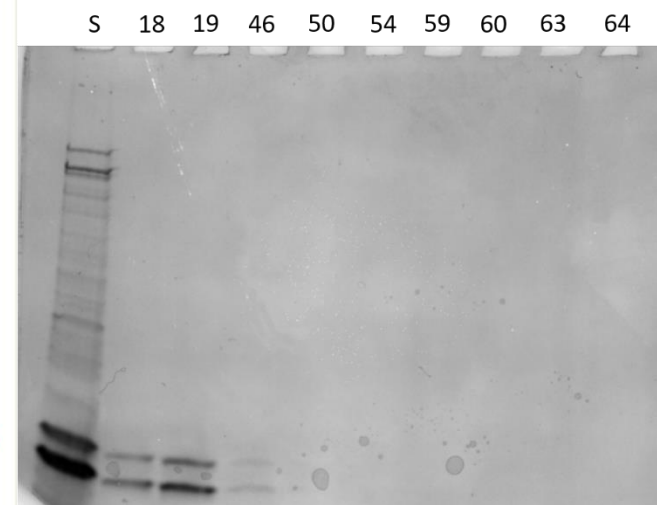
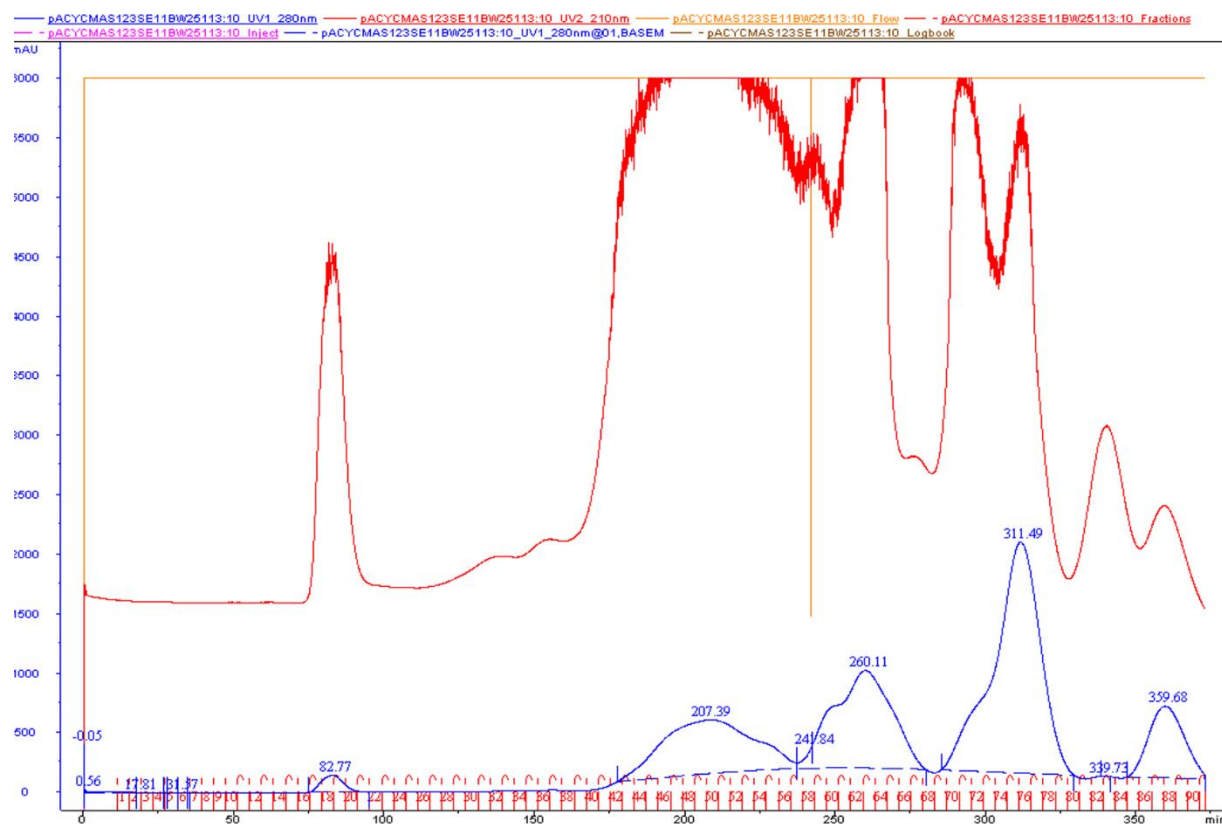


Figure 6.13: Sephacryl S300 purification of *E. coli* SE11 subunits/polymer. Chromatogram and SDS-PAGE of pACYCD1-MA-S123. Peak fractions were boiled for 15 mins at $>95^{\circ}\text{C}$ in SDS-PAGE loading dye prior to SDS-PAGE. Number correspond to eluted fraction number of 1.5 ml each fractions. S indicates sample before purification. *Spill over of sample 19 was observed at lane 46.*

6.5 Conclusions

In this chapter, the *E. coli* SE11 p000001-p00005 CU fimbriae operon from pSE11-2 plasmid has been analysed by bioinformatics analysis and identity of the subunits validated by *in vitro* analysis. The operon consists of an usher with 70% identity with Caf1A, a chaperone with 59% identity with Caf1M and 2 subunits, named subunit 1 and subunit 2 with 34% and 31% identity to Caf1, respectively. A third ORF also annotated as subunit, appears too small to be a functional subunit. Recombinant constructs successfully produced surface assembled polymer of subunits 1 and 2, and a separate construct produced low levels of periplasmic subunit and chaperone. Mass spectrometry was used to confirm identity of both subunits. Attempts to purify subunit:chaperone complex using a C-terminally His tagged chaperone were unsuccessful. However, polymer preparations with very high levels of subunit1 and 2 were successfully purified from *E. coli* BL21/pACYCD1-MA-S123 by heat extraction, Triton TX114 extraction and size exclusion chromatography. Following confirmation of purity, this preparation can be used for antibody production for further study, including testing any cross- reaction with F1 polymer.

Identity between F1 and the *E. coli* SE11 polymers is not particularly high, although B cell epitope prediction revealed that loop 5, which contains the conserved Gly Asn sequence in both, was predicted as a discontinuous B-Cell epitope. The ancestral origin of the *caf* operon and the link between this and the closely related *E. coli* SE11 operon is also of interest. A plasmid comparison between *Y. pestis* (pMT1 and pFra), *E. coli* SE11 (ECSE_P2) and *S. enterica* (pHCM2) showed that despite the higher similarity between pMT1 and pFra with pHCM2 (56.2% and 57.6% respectively), pHCM2 does not have any

similarity with the Caf operon. In contrast, despite the lower similarity between pMT1 and pFra with ECSE_P2, there is strong similarity between the 2 chaperone usher loci. Unfortunately, the creation of a hybrid *E. coli* SE11 *caf* CU system to determine the flexibility of the usher to export Caf1 polymer was unable to be performed due to time constraints.

**CHAPTER 7 GENERAL
DISCUSSION AND FUTURE
DIRECTIONS**

7.1 Properties of the Caf1A usher

In this study, modelling of the Caf translocon was used to identify potential residues that might be important in polymerisation and translocation of Caf1. The first step of F1 fibre translocation occurs when the Caf1M:Caf1 complex binds to the N-terminus of Caf1A and opens the pore (Sauer et al., 2002; Yu et al., 2012). In PapC, a single salt bridge (Arg305 in the plug interacts with Glu467 in an α -helix within the barrel) was shown to keep the plug in the closed position (Volkan et al., 2013). Arg305 corresponds to Glu291 in Caf1A. In the closed Caf translocon, Glu291 is shown to interact with Lys462 in the β -barrel. The mutations E291K, E291A and K462A all resulted in an apparent increase in the level of surface F1. This would be consistent with the mutation weakening the ionic interaction and facilitating opening of the pore and F1 export. Of the usher residues mutated in this study, the one that had the highest impact on the export of surface F1 when mutated was Ser289, which corresponds to R303 in PapC. Ser289 was shown in the close Caf1A model to interact with a residue in β -hairpin, Asp217. The profound reduction of surface F1 with the Caf1 S289A mutation suggests that this residue could be important in translocating the polymer through the pore by direct interaction with Caf1 subunit or through an additional interaction in the pore. Tyr266 was also shown to be important in positioning the plug inside the pore. In the closed Caf1A model, Tyr266 formed an extensive hydrophobic interaction with Tyr198 (β 5-6 hairpin) and Tyr215 (β 5-6 hairpin). In PapC, the β 5-6 hairpin was also shown to be important in stabilising the plug inside PapC β -barrel as deletion of this region significantly increased pore permeability in an Erythromycin sensitivity assay (Volkan et al., 2013; Mapingire et al., 2009). Much more detailed investigation is required to establish involvement of these three Tyr residues in pore gating.

7.2 Caf1 as scaffold of short peptides

A total of four loops were replaced with Gly and the charged residues (ELDKWA) to test the ability of Caf1 to act as scaffold for surface epitope. The modelled Caf translocon was used to explain problems in constructs that appeared to be non-permissive and predict mutations within the usher that might improve export. The most promising site among these four loops for insertion of short foreign epitope is loop 5. Insertion of Gly sequence or a charged epitope to this loop retained a significant level of surface F1, although there was a large reduction with the charged loop. The translocon model was used to repair the clash between the inserted charged residues with Glu169 in the β -barrel. This improved export from only $37.41\% \pm 1.26$ of surface F1 from cells carrying pACYCF1::1 to $78.71\% \pm 1.47$ and $86.54\% \pm 3.86$ from cells carrying pACYCF1::1_Q169A and pACYCF1::1_Q169G respectively. Focusing on one identified permissive site loop 5, this loop was used to insert polyhistidine-tag for visualisation of F1 fibres using NTA-fluorescent probes. Analysis of F1-His polymers expressed on the surface showed that Caf1, with 4his tagged can be assembled and efficiently polymerise. but that cells expressing a 6his-tagged insertion in loop 5 displayed signs of toxicity in the cells and would be problematic for assessing binding of NTA-fluorescent probe.

7.3 *E. coli* SE11 fimbriae operon similar to *caf* operon

Investigation of closely related ushers with Caf1A revealed that the usher from *E. coli* SE11 fimbriae operon from pSE11-2 shares 70% amino acid sequence identity over the whole protein. This newly identified usher was shown to group with $\gamma 3$ fimbriae and next to Caf1A. Locus analysis showed that this operon consists of one chaperone, one usher, and

three subunits. However, subunit 3 is too small for a fimbrial subunit and cloning of this operon into pACYCDuet1 and pTRC99, and mass spectrometry identified only two expressed subunits, subunits 1 and 2. How these subunits assemble, within a single fibre or separate fibres is an interesting question for further study. This relates to other γ 3-fimbriae members as the operon consists of one usher, one chaperone and one or two subunits. The subunit(s) assembled into thin and flexible fimbriae. The pTRC99-M6his-S1 and pTRC99-M8His-S123 were used to purify the subunits but were unsuccessful due to a low level of these subunits accumulated in the periplasm. Therefore, the subunits were purified following heat extraction from the cell surface and detergent extraction, by size exclusion from the pACYCD1-MA-S123 construct. The purified protein will be used for antibody production which can then be used to for further study of these fibres as well as testing/excluding any cross-reaction with F1 polymer.

7.4 Future directions

Key future work is to assess any cross reaction between antibody against *E. coli* SE11 subunits with F1 and Caf1. ELISA and fluorescent microscopy would both be suitable to determine any cross reaction. The flexibility of usher to export other subunits by infusing the *E. coli* SE11 usher into *caf* operon by replacing Caf1A can also be determined. The epifluorescence assay for binding of polyhistidine tagged Caf1 to a fluorescent probe should be established, as Caf1 inserted with 4His was successfully exported as polymer to the cell surface. Direct NTA-fluorescein binding to F1 polymer could be an efficient quantitative method for assembly of F1. For this, F1 polymers carrying polyhistidine-tag should assemble efficiently and be non-toxic. Another outstanding question to be answered is what

is the function of the hydrophobic path (Tyr). These residues must be mutated to identify the function of these tyrosines. Might these Tyrosine residues involve in gating the pore?

Appendix 1: Usher proteins from similarity search results, using BLASTP, used to construct phylogenetic tree.

Proteins not previously included in earlier γ fimbrial clade are highlighted in yellow. Duplicates (100% identical) have been removed.

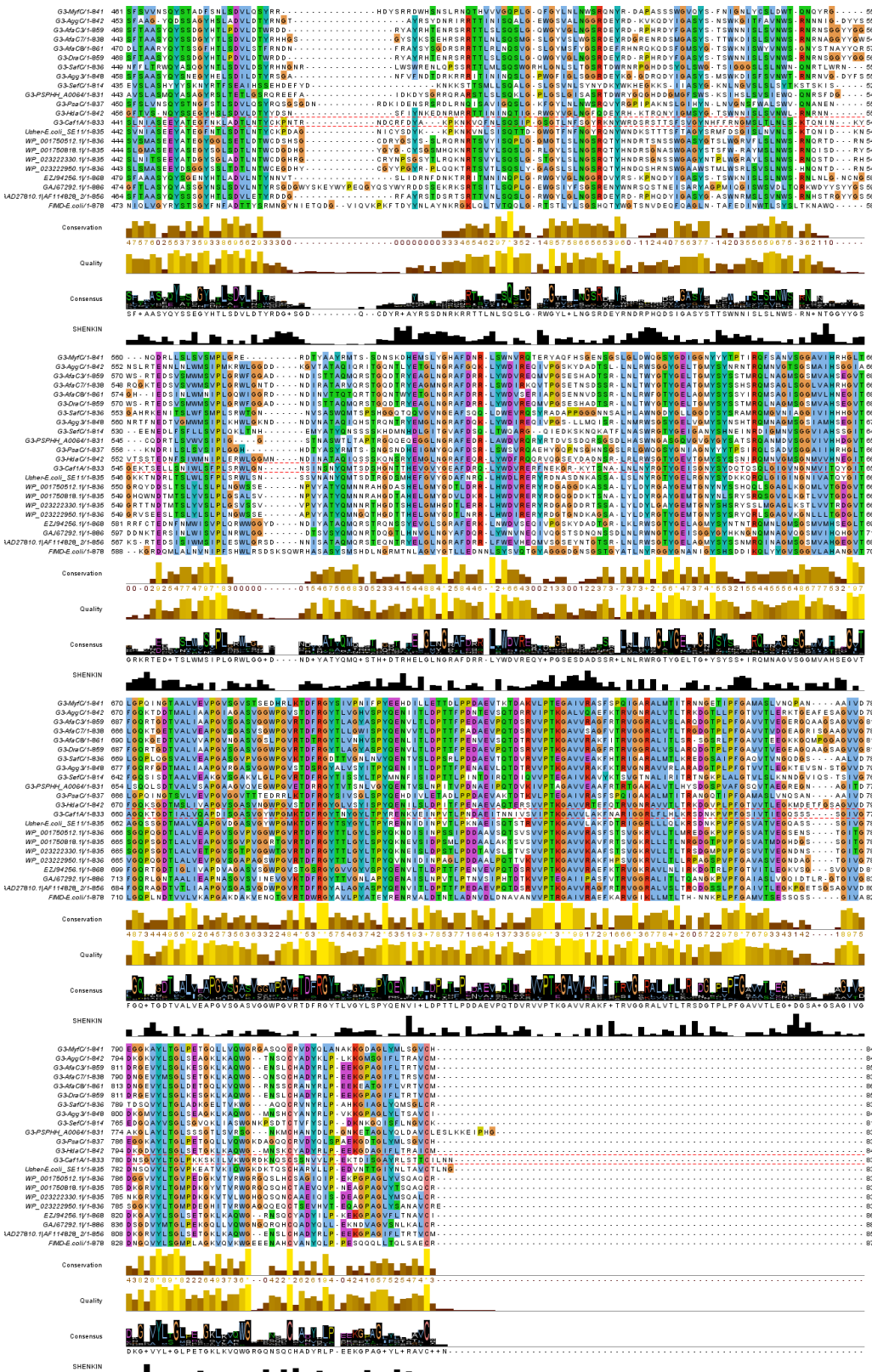
Description	Query cover	E value	Ident	Accession
fimbrial usher protein [<i>Yersinia pestis</i>]	100%	0	100%	WP_002211763.1
F1 capsule-anchoring protein [<i>Escherichia coli</i> SE11]	99%	0	70%	WP_000877656.1
putative fimbrial usher protein [<i>Salmonella enterica</i>]	99%	0	49%	WP_001750512.1
putative fimbrial usher protein [<i>Salmonella enterica</i>]	99%	0	47%	WP_023222950.1
putative fimbrial usher protein [<i>Salmonella enterica</i>]	99%	0	49%	WP_001750818.1
putative fimbrial usher protein [<i>Salmonella enterica</i>]	99%	0	47%	WP_023222330.1
putative fimbrial usher protein [<i>Salmonella enterica</i>]	99%	0	47%	WP_023248754.1
outer membrane usher protein AfaC [<i>Escherichia coli</i>]	99%	0	44%	WP_021573024.1
putative outer membrane usher protein psac [<i>Escherichia coli</i>]	99%	0	44%	WP_023141262.1
pilin outer membrane usher protein SafC [<i>Escherichia coli</i>]	99%	0	44%	WP_001205747.1
Outer membrane usher protein fimD precursor [<i>Escherichia coli</i> ISC41]	99%	0	44%	CDL44533.1
F1 capsule-anchoring protein [<i>Escherichia coli</i> 1-110-08_S4_C1]	99%	0	44%	EYE06634.1
pilin outer membrane usher protein SafC [<i>Escherichia coli</i>]	99%	0	44%	WP_001330422.1
pilin outer membrane usher protein SafC [<i>Salmonella enterica</i>]	99%	0	44%	WP_023219945.1
Outer membrane usher protein SfmD [<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>Johannesburg</i> str. S5-703]	99%	0	44%	EHC69713.1
DraC [<i>Escherichia coli</i>]	99%	0	44%	AAK16477.1
DafaC [<i>Escherichia coli</i>]	99%	0	44%	AAK17181.1
pilin outer membrane usher protein SafC [<i>Edwardsiella hoshinae</i>]	99%	0	44%	WP_024524433.1
pilin outer membrane usher protein SafC [<i>Yersinia ruckeri</i>]	99%	0	44%	WP_004719950.1
fimbrial usher protein [<i>Edwardsiella piscicida</i>]	99%	0	44%	GAJ67292.1
Outer membrane usher protein SfmD [<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>Minnesota</i> str. A4-603]	99%	0	44%	EHC74323.1

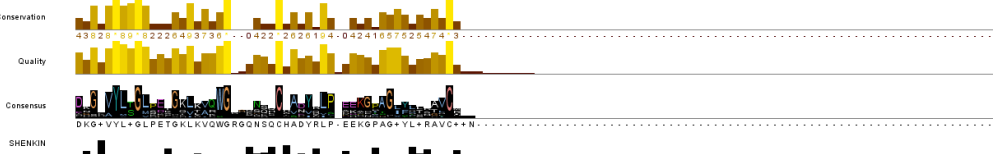
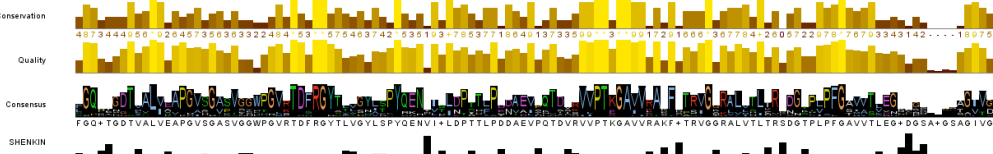
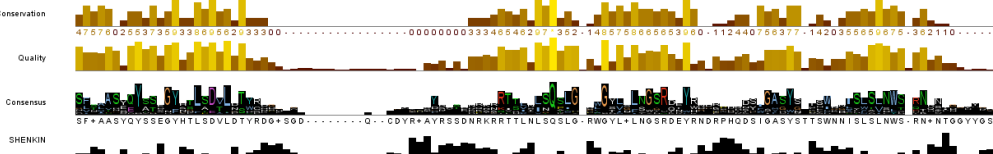
aggregative adherence fimbria II usher protein AafC [Escherichia coli O42]	99%	0	43%	AAD27810.1
HdaC, HUS-associated diffuse adherence [Salmonella enterica]	99%	0	42%	WP_023221224.1
type VII secretion system (T7SS), usher family protein [Escherichia coli 2-005-03_S3_C2]	99%	0	41%	KDA74785.1
type VII secretion system (T7SS), usher family protein [Escherichia coli 2-005-03_S1_C3]	99%	0	41%	EZK11614.1
type VII secretion system (T7SS), usher family protein [Escherichia coli 1-250-04_S1_C3]	99%	0	41%	EZJ94256.1
Outer membrane usher protein Agg3C [Escherichia coli]	99%	0	41%	WP_001403380.1
outer membrane usher protein AggC [Escherichia coli]	99%	0	41%	WP_021523943.1
P pilus assembly protein, porin PapC [Serratia marcescens]	99%	0	42%	WP_015673850.1
hypothetical protein [Providencia stuartii]	99%	0	39%	WP_004927466.1
psaC [Yersinia pestis EV76-CN]	99%	0	39%	CAA66357.1
fimbrial outer membrane usher protein SthC [Serratia odorifera]	98%	0	39%	WP_004957477.1
fimbrial outer membrane usher protein SthC [Enterobacter mori]	98%	0	38%	WP_010431832.1
fimbrial outer membrane usher protein SthC [Cronobacter sakazakii]	97%	0	38%	WP_012124895.1
fimbrial outer membrane usher protein SthC [Klebsiella oxytoca]	97%	0	38%	WP_004851752.1
type 1 fimbriae anchoring protein FimD [Klebsiella oxytoca HKOPL1]	97%	0	38%	AHW88442.1
outer membrane usher protein LpfC [Enterobacter cloacae]	99%	0	37%	WP_023303537.1
outer membrane usher protein FimD [Enterobacter hormaechei]	99%	7.00E-180	37%	WP_017693574.1
fimbrial outer membrane usher protein SthC [Enterobacter cloacae]	99%	9.00E-180	37%	WP_017384637.1
type 1 fimbriae anchoring protein FimD [Serratia fonticola]	99%	4.00E-178	38%	WP_021806318.1
fimbrial outer membrane usher protein SthC [Yersinia aldovae]	98%	5.00E-174	37%	WP_004700856.1
outer membrane usher protein SfmD [Serratia marcescens] gamma1	99%	1.00E-169	36%	WP_004935286.1
fimbrial outer membrane usher protein SthC [Klebsiella oxytoca]	99%	2.00E-169	37%	WP_004101585.1
MULTISPECIES: fimbrial assembly protein [Serratia]	96%	2.00E-169	37%	WP_025122548.1
putative uncharacterized protein [Klebsiella variicola CAG:634]	97%	3.00E-169	38%	WP_022064652.1
fimbrial outer membrane usher protein SthC [Klebsiella oxytoca]	99%	5.00E-169	37%	WP_016807753.1

fimbrial outer membrane usher protein SthC [Yersinia enterocolitica]	99%	1.00E-168	37%	WP_005165265.1
P pilus assembly protein, porin PapC [Serratia marcescens]	95%	4.00E-168	39%	WP_015673631.1
outer membrane usher protein SfmD [Salmonella enterica]	98%	4.00E-167	37%	WP_023136615.1
fimbrial outer membrane usher protein SthC [Salmonella enterica]	98%	7.00E-167	37%	WP_000990588.1
type 1 fimbriae anchoring protein FimD [Klebsiella pneumoniae]	97%	1.00E-165	37%	WP_016531172.1
fimbrial outer membrane usher protein SthC [Klebsiella pneumoniae]	99%	8.00E-165	36%	WP_011977702.1
fimbrial outer membrane usher protein StbC [Salmonella enterica]	98%	2.00E-164	36%	WP_000300173.1
type 1 fimbriae anchoring protein FimD [Serratia fonticola]	98%	4.00E-164	37%	WP_021806158.1
putative outer membrane usher protein psac [Klebsiella pneumoniae]	99%	1.00E-163	36%	WP_023158088.1
outer membrane usher protein LpfC [Klebsiella oxytoca]	94%	2.00E-162	36%	WP_004132915.1
fimbrial outer membrane usher protein SthC [Serratia plymuthica]	98%	3.00E-162	37%	WP_006318902.1
outer membrane usher protein LpfC [Enterobacter sp. BIDMC 29]	99%	4.00E-162	36%	EUM12550.1
Outer membrane usher protein LpfC [Klebsiella sp.]	99%	5.00E-162	36%	CDQ17421.1
Outer membrane usher protein LpfC precursor [Enterobacter sp. DC3]	99%	5.00E-162	36%	EWG68909.1
type 1 fimbriae anchoring protein FimD [Klebsiella pneumoniae IS39]	92%	2.00E-161	38%	CDL61039.1
outer membrane usher protein LpfC [Enterobacter sp. MGH 38]	99%	8.00E-161	36%	WP_023324354.1
hypothetical protein L465_04515 [Enterobacter sp. BIDMC 29]	99%	8.00E-161	36%	EUM07318.1
outer membrane usher protein LpfC [Enterobacter sp. MGH 14]	99%	9.00E-161	36%	WP_022650650.1
type 1 fimbriae anchoring protein FimD [Enterobacter cloacae ISC8]	99%	1.00E-160	36%	CDL32450.1
fimbrial outer membrane usher protein SthC [Klebsiella pneumoniae]	99%	2.00E-160	36%	WP_004177742.1
putative outer membrane usher protein psac [Klebsiella pneumoniae]	99%	3.00E-159	35%	WP_023157904.1
fimbrial assembly protein [Enterobacter asburiae]	99%	1.00E-158	36%	WP_024908720.1
outer membrane usher protein LpfC [Enterobacter sp. MGH 24]	99%	1.00E-158	36%	WP_023334990.1
outer membrane usher protein [Klebsiella pneumoniae]	99%	1.00E-158	35%	WP_023296915.1
outer membrane usher protein LpfC [Enterobacter cloacae]	99%	2.00E-158	35%	WP_023299644.1
outer membrane usher protein FimD [Enterobacter hormaechei]	99%	8.00E-158	35%	WP_017693768.1

type 1 fimbriae anchoring protein FimD [Cronobacter dublinensis]	99%	1.00E-157	37%	WP_007756586.1
fimbrial outer membrane usher protein SthC [Providencia rustigianii]	99%	3.00E-157	35%	WP_006813805.1
Outer membrane usher protein [Klebsiella pneumoniae BJ1-GA]	97%	5.00E-157	36%	CDI26527.1
fimbrial outer membrane usher protein SthC [Enterobacter cancerogenus]	98%	6.00E-157	36%	WP_006174891.1
chaperone-usher secretion system usher protein [Klebsiella sp. AS10]	99%	6.00E-157	36%	EUB41790.1
putative fimbrial outer membrane usher protein SthC [Salmonella enterica]	97%	8.00E-157	36%	WP_001727856.1
fimbrial outer membrane usher protein SthC [Klebsiella sp. OBRC7]	99%	1.00E-156	36%	WP_009652085.1
P pilus assembly protein, porin PapC [Enterobacter cloacae]	99%	1.00E-156	35%	WP_015571070.1
fimbrial outer membrane usher protein SthC [Salmonella enterica]	97%	1.00E-156	36%	WP_001536549.1
outer membrane usher protein LpfC [Enterobacter sp. MGH 33]	99%	1.00E-156	35%	EUM54096.1
putative outer membrane usher protein psac [Salmonella enterica]	97%	2.00E-156	36%	WP_001751132.1
putative fimbrial outer membrane usher protein SthC [Salmonella enterica]	97%	2.00E-156	36%	WP_023197800.1
outer membrane usher protein [Salmonella enterica]	97%	2.00E-156	36%	WP_023138573.1
fimbrial outer membrane usher protein SthC [Salmonella enterica]	97%	2.00E-156	36%	WP_006495791.1
Fimbriae usher protein StcC [Salmonella enterica subsp. enterica serovar Inverness str. R8-3668]	97%	3.00E-156	36%	EHC63580.1
fimbrial assembly protein [Salmonella enterica]	97%	3.00E-156	35%	WP_024134916.1
fimbrial outer membrane usher protein SthC [Salmonella enterica]	97%	3.00E-156	36%	WP_001584422.1
fimbrial outer membrane usher protein SthC [Klebsiella oxytoca]	99%	3.00E-156	36%	WP_014838686.1
fimbrial outer membrane usher protein SthC [Salmonella enterica]	97%	3.00E-156	36%	WP_001160702.1
Fimbriae usher protein StcC [Salmonella enterica subsp. enterica serovar Hvittingfoss str. A4-620]	97%	3.00E-156	36%	EHC42566.1
fimbrial outer membrane usher protein SthC [Salmonella enterica]	97%	4.00E-156	36%	WP_001551945.1
fimbrial outer membrane usher protein SthC [Salmonella enterica]	97%	4.00E-156	36%	WP_001160708.1
putative fimbrial outer membrane usher protein SthC [Salmonella enterica]	97%	4.00E-156	36%	WP_023203387.1
fimbrial outer membrane usher protein SthC [Klebsiella pneumoniae]	97%	4.00E-156	36%	WP_022631368.1
putative fimbrial outer membrane usher protein SthC [Salmonella enterica]	97%	4.00E-156	36%	WP_023233222.1

Appendix 2: Multiple sequence alignment of all $\gamma 3$ fimbrial.





Appendix 3: Impact on subunit polymerization using plasmid pFM1 and surface polymer using plasmid pFMA1.

Caf1 residues removed	Inserted amino acids sequence	Level of protein recovery	Polymerization (pFM1)	α GP41 mAb/ α F1 polymer	Surface polymer (pFMA1) SDS_PAGE	Surface immunofluorescence ($\times 10^3$) α GP41mAb/ anti-F1 polyclonal
D33N34	F1::3 (ELDKWA)	+++	1-8 subunits	+++ / ++	VHMW	0.82 \pm 0.08 / 109 \pm 1.5
	F1D11 (GG)	++++	HMW (sim to WT)	++	VHMW	
S76QDGNNH82	F1::2 (ELDKWA)	+	Regular polymer, faint	Nd / ++	VHMW	Nd / 54.3 \pm 8.4
	F1D12 (GGGG)	++++	HMW (identical to WT)	++	VHMW**	
D92SRDF96	F1::1 (ELDKWA)	++	HMW (less than WT)	+(faint) / +++	VHMW, some in well	0.92 \pm 0.27 / 192 \pm 10.9
	F1D13 (GGGG)	++++	HMW (sim to WT)	++	None** exported	
K129GGK132	F1::4 (ELDKWA)	++	1-8 subunits	++ / ++	Aggregate in well	2.81 \pm 0.25 / 52.1 \pm 0.7
	F1D14 (DGGG)	++++	HMW (possibly less, banding differently)	++	VHMW	
None	WT	++++	HMW	- / +++++	VHMW	0.29 \pm 0.04 / 269 \pm 23.9

Available previous unpublished results from this laboratory, Elham Moslehi and Snezana Vasiljec. Plasmid pFM1 encodes chaperone and subunit only and results in high levels of periplasmic polymer. pFMA1 also encodes usher which leads to surface assembly of F1, but only at low levels. Table above shows that polymerisation of F1D11, F1D12 and F1D13 were similar to wildtype and F1D14 also polymerises but with a slightly shifted size. All of these four loop mutants produced high molecular weight polymer. F1::1, also produces high molecular weight in the periplasm although less than WT F1; F1::2 has low level of polymer with regular faint polymer; F1::3 and F1::4 accumulate oligomers of around 1-8 subunits suggesting a problem with polymerisation. There was an indication of surface polymer in most constructs, but this plasmid produces lower levels. **For the surface polymer in pFMA1 (chaperone, subunit and usher), it has been concluded that the results of F1D13 and F1D12 must have been mixed up and should be the other way around.

Appendix 4: Caf1 sequence with G or ELDKWA insertion in FASTA format.
Coloured in red and italic indicates replacement

Loop mutation	Sequence	Description
Caf1-D11 (D33N34)	>F1D11 ADLTASTTATATLVEPARITLTYKEGAPITIMGGGNIDTELLVGTLLTGG YKTGTTSTSVNFTDAAGDPMYLTFTSQDGNNHQFTTKVIGKDSRDFDISP KVNGENLVGDDVVLATGSQDFFVRSIGSKGGKLAAGKYTDAVTVTVSNQ	Replacement
Caf1-D12 (S76QDGNNH8 2)	>F1D12 ADLTASTTATATLVEPARITLTYKEGAPITIMDNGNIDTELLVGTLLTGG YKTGTTSTSVNFTDAAGDPMYLTFTGGGQFTTKVIGKDSRDFDISPKVN GENLVGDDVVLATGSQDFFVRSIGSKGGKLAAGKYTDAVTVTVSNQ	Reduction from 7 to 4 amino acids
Caf1-D13 (D92SRDF96)	>F1D13 ADLTASTTATATLVEPARITLTYKEGAPITIMDNGNIDTELLVGTLLTGG YKTGTTSTSVNFTDAAGDPMYLTFTSQDGNNHQFTTKVIGKGGGDISPK VNGENLVGDDVVLATGSQDFFVRSIGSKGGKLAAGKYTDAVTVTVSNQ	Reduction from 5 to 4 amino acids
Caf1-D14 (K129GGK132)	>F1D14 ADLTASTTATATLVEPARITLTYKEGAPITIMDNGNIDTELLVGTLLTGG YKTGTTSTSVNFTDAAGDPMYLTFTSQDGNNHQFTTKVIGKDSRDFDISP KVNGENLVGDDVVLATGSQDFFVRSIGSDGGGLAAGKYTDAVTVTVSNQ	Replacement
Caf1::3 (D33N34)	>F1::3 ADLTASTTATATLVEPARITLTYKEGAPITIMELDKWAGNIDTELLVGTLL TGGYKTGTTSTSVNFTDAAGDPMYLTFTSQDGNNHQFTTKVIGKDSRDF DISPKVNGENLVGDDVVLATGSQDFFVRSIGSKGGKLAAGKYTDAVTVTV SNQ	Converted from 5 to 6 amino acids
Caf1::2 (S76QDGNNH8 2)	> F1::2 ADLTASTTATATLVEPARITLTYKEGAPITIMDNGNIDTELLVGTLLTGG YKTGTTSTSVNFTDAAGDPMYLTFTELDKWAQFTTKVIGKDSRDFDISPK VNGENLVGDDVVLATGSQDFFVRSIGSKGGKLAAGKYTDAVTVTVSNQ	Replacement
Caf1::1 (D92SRDF96)	> F1::1 ADLTASTTATATLVEPARITLTYKEGAPITIMDNGNIDTELLVGTLLTGG YKTGTTSTSVNFTDAAGDPMYLTFTSQDGNNHQFTTKVIGKELDKWADIS PKVNGENLVGDDVVLATGSQDFFVRSIGSKGGKLAAGKYTDAVTVTVSNQ	Converted from 2 to 6 amino acids
Caf1::4 (K129GGK132)	>F1::4 ADLTASTTATATLVEPARITLTYKEGAPITIMDNGNIDTELLVGTLLTGG YKTGTTSTSVNFTDAAGDPMYLTFTSQDGNNHQFTTKVIGKDSRDFDISP KVNGENLVGDDVVLATGSQDFFVRSIGSELDKWALAAGKYTDAVTVTVSN Q	Converted from 4 to 6 amino acids

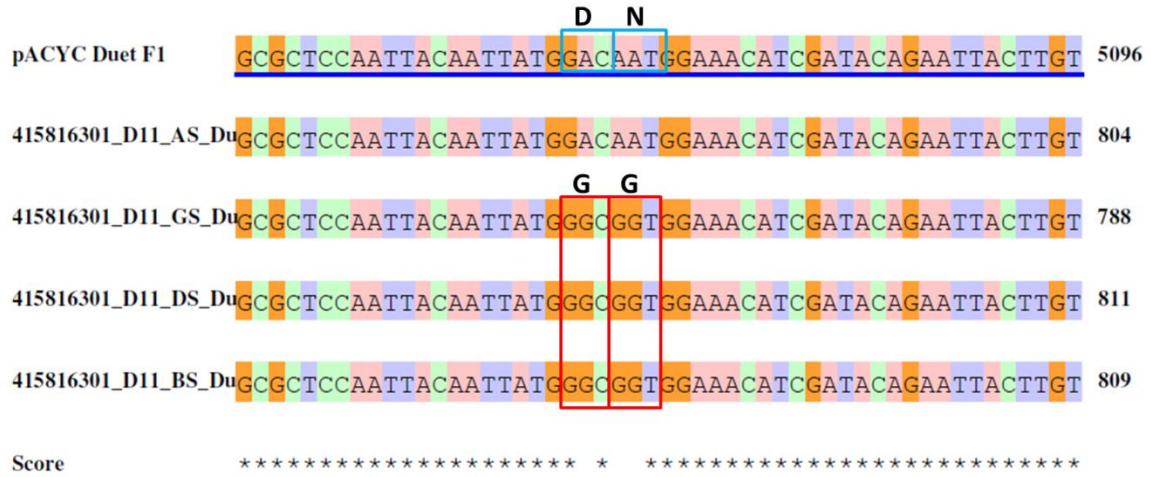
Appendix 5: Alignment of pACYC-F1D13 sequencing results.

Alignment of pACYC-F1D13 constructs using DNAdynamo. Cyan box indicates residues in WT and red box indicates the changed residues in mutant. From the alignment, only F1D13-E (D13_5S) from Stellar confirmed with loop mutant.

pACYC Duet F1	CTCAGGATGGAAATAACCAATTCCTACAAAAGTGATTGGCAAGGATTC	5246
415542001_d13_1S_Du	CTCAGGATGGAAATAACCAATTCCTACAAAAGTGATTGGCAAGGATTC	958
415542001_d13_2S_Du	CTCAGGATGGAAATAACCAATTCCTACAAAAGTGATTGGCAAGGATTC	954
415542001_d13_3S_Du	CTCAGGATGGAAATAACCAATTCCTACAAAAGTGATTGGCAAGGATTC	956
415542001_d13_5S_Du	CTCAGGATGGAAATAACCAATTCCTACAAAAGTGATTGGCAAGGATTC	963
415542001_d13_1T_Du	CTCAGGATGGAAATAACCAATTCCTACAAAAGTGATTGGCAAGGATTC	958
415542001_d13_2T_Du	CTCAGGATGGAAATAACCAATTCCTACAAAAGTGATTGGCAAGGATTC	967
415542001_d13_3T_Du	CTCAGGATGGAAATAACCAATTCCTACAAAAGTGATTGGCAAGGATTC	945
415542001_d13_5t_Du	CTCAGGATGGAAATAACCAATTCCTACAAAAGTGATTGGCAAGGATTC	938
415542001_d13_7T_Du	CTCAGGATGGAAATAACCAATTCCTACAAAAGTGATTGGCAAGGATTC	949
Score	*****	
pACYC Duet F1	TAGAGATTTTGATATCTCTCCTAAGGTAAACGGTGAGAACCTTGTGGGGGAT	5298
415542001_d13_1S_Du	TAGAGATTTTGATATCTCTCCTAAGGTAAACGGTGAGAACCTTGTGGGGGAT	1010
415542001_d13_2S_Du	TAGAGATTTTGATATCTCTCCTAAGGTAAACGGTGAGAACCTTGTGGGGGAT	1006
415542001_d13_3S_Du	TAGAGATTTTGATATCTCTCCTAAGGTAAACGGTGAGAACCTTGTGGGGGAT	1008
415542001_d13_5S_Du	CGGAGGT---GATATCTCTCCTAAGGTAAACGGTGAGAACCTTGTGGGGGAT	1012
415542001_d13_1T_Du	TAGAGATTTTGATATCTCTCCTAAGGTAAACGGTGAGAACCTTGTGGGGGAT	1010
415542001_d13_2T_Du	TAGAGATTTTGATATCTCTCCTAAGGTAAACGGTGAGAACCTTGTGGGGGAT	1019
415542001_d13_3T_Du	TAGAGATTTTGATATCTCTCCTAAGGTAAACGGTGAGAACCTTGTGGGGGAT	997
415542001_d13_5t_Du	TAGAGATTTTGATATCTCTCCTAAGGTAAACGGTGAGAACCTTGTGGGGGAT	990
415542001_d13_7T_Du	TAGAGATTTTGATATCTCTCCTAAGGTAAACGGTGAGAACCTTGTGGGGGAT	1001
Score	*** *	

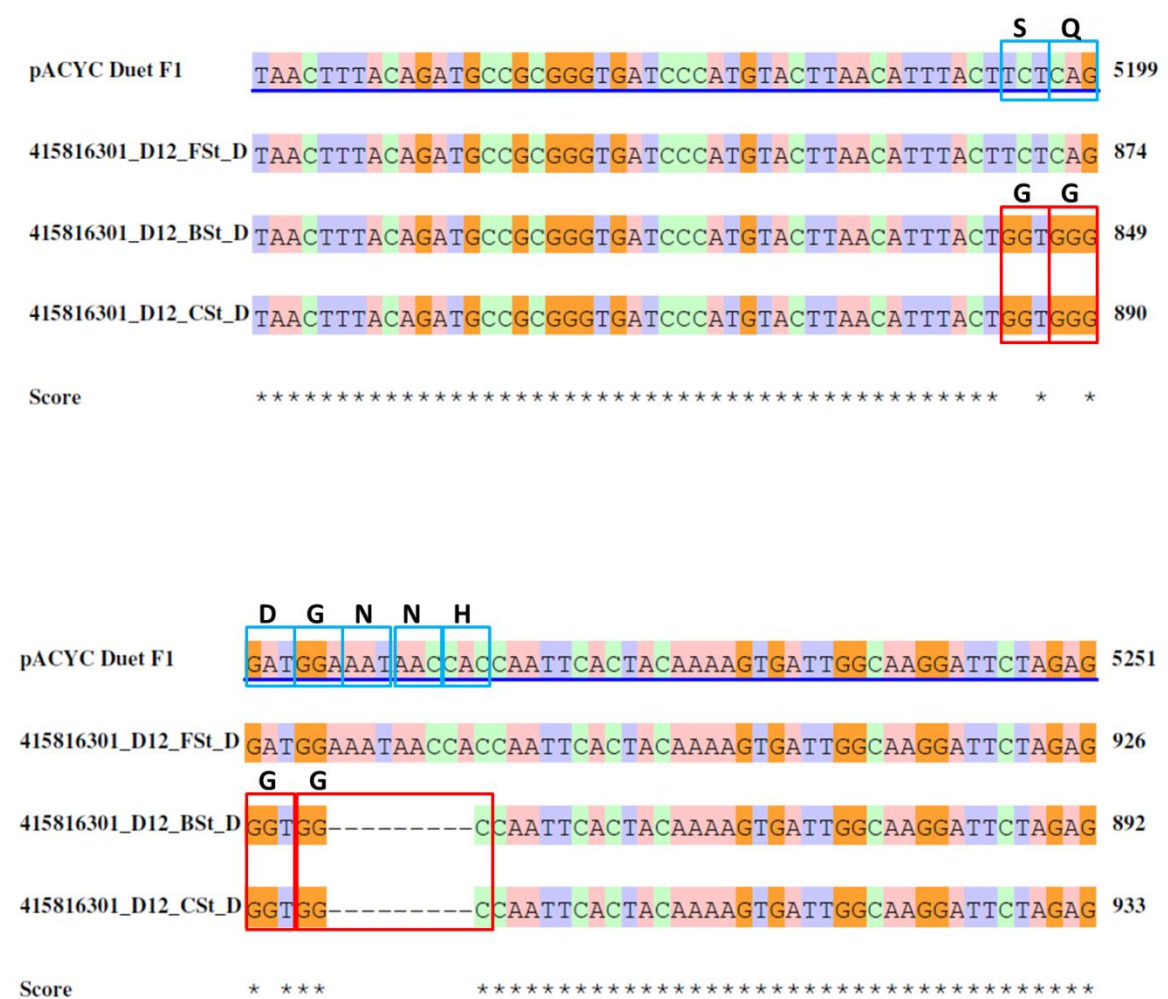
Appendix 6: Alignment of pACYC-F1D11 sequencing results

Alignment of pACYC-F1D11 constructs using DNAdynamo. Cyan box indicates residues in WT and red box indicates the changed residues in mutant. From the alignment, only F1D11-B, -D and -G from Stellar cells confirmed with loop mutant.



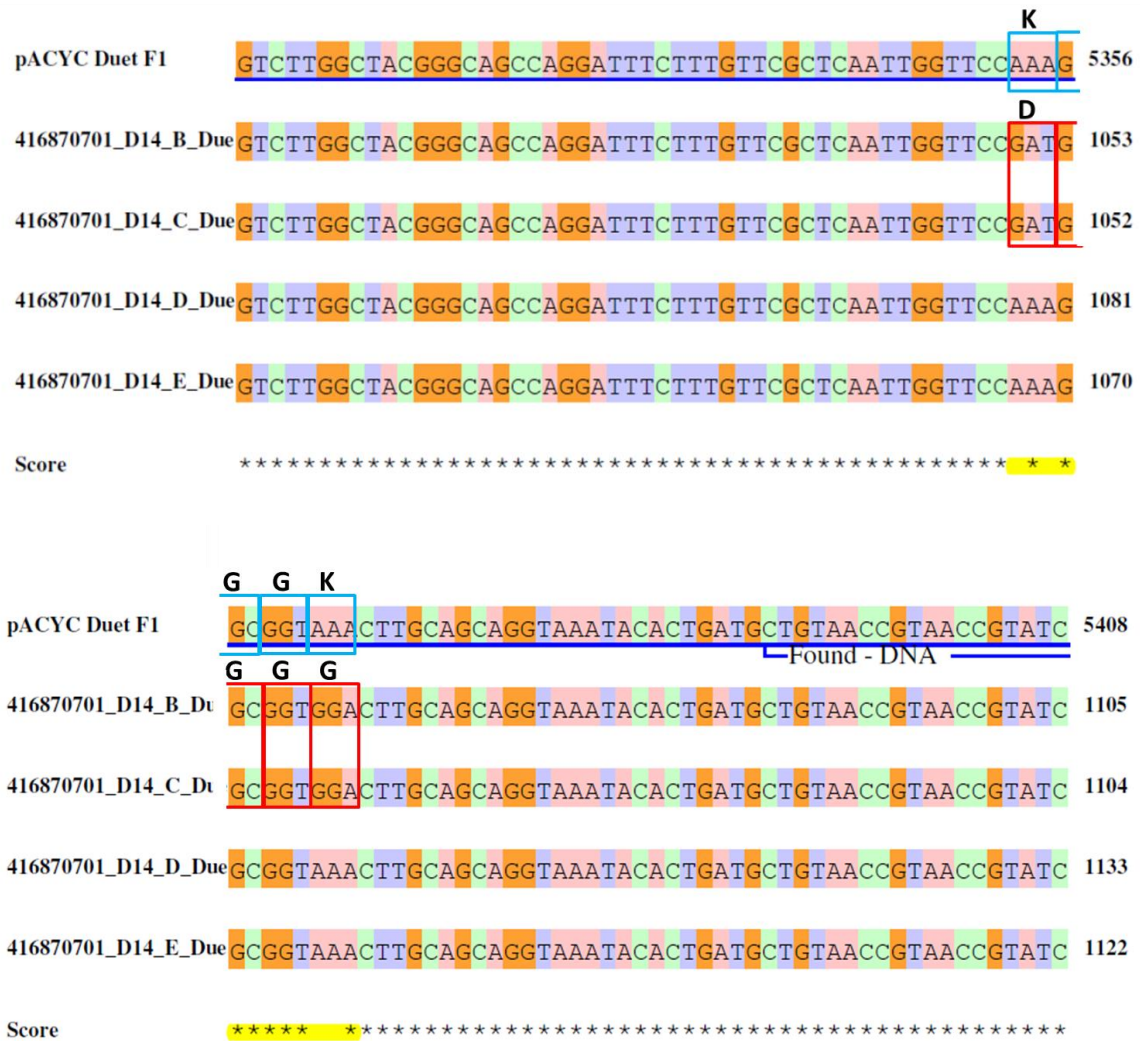
Appendix 7: Alignment of pACYC-F1D12 sequencing results.

Alignment of pACYC-F1D12 constructs using DNAdynamo. Cyan box indicates residues in WT and red box indicates the changed residues in mutant. From the alignment, only F1D12-B and -C from Stellar cells confirmed with loop mutant.



Appendix 8: Alignment of pACYC-F1D14 sequencing results.

Alignment of pACYC-F1D14 constructs using DNAdynamo. Cyan box indicates residues in WT and red box indicates the changed residues in mutant. From the alignment, only F1D14-B and -C from Stellar cells confirmed with loop mutant.



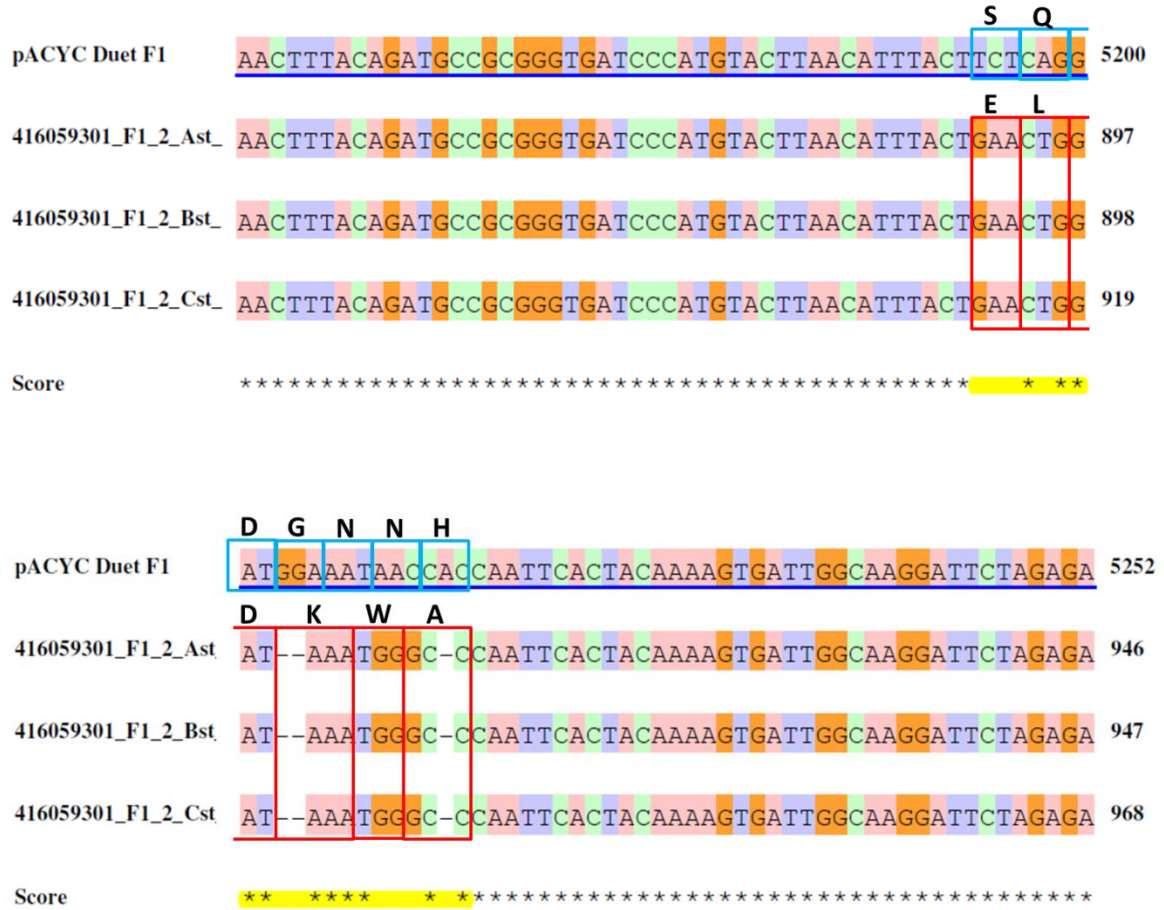
Appendix 9: Alignment of pACYC-F1::1 sequencing results.

Alignment of pACYC-F1::1 constructs using DNAdynamo. Cyan box indicates residues in WT and red box indicates the changed residues in mutant. From the alignment, only F1::1-C from Stellar cells confirmed with loop mutant.

pACYC Duet F1	ATGGAAATAACCACCAATTCACTACAAAAGTGATTGGCAAGGATTC---TAG	5249
	<div>D S R</div>	
416870701_Ins1_B_Du	ATGGAAATAACCACCAATTCACTACAAAAGTGATTGGCAAGGATTC---TAG	964
416870701_Ins1_C_Du	ATGGAAATAACCACCAATTCACTACAAAAGTGATTGGCAAGGAACTGGATAA	924
	<div>E L D K</div>	
416870701_Ins1_D_Du	ATGGAAATAACCACCAATTCACTACAAAAGTGATTGGCAAGGATTC---TAG	928
416870701_Ins1_E_Du	ATGGAAATAACCACCAATTCACTACAAAAGTGATTGGCAAGGATTC---TAG	936
Score	*****	**
pACYC Duet F1	AGATTTTGATATCTCTCCTAAGGTAAACGGTGAGAACCTTGTGGGGGATGAC	5301
	<div>D F</div>	
	*misc_difference call	
416870701_Ins1_B_Du	AGATTTTGATATCTCTCCTAAGGTAAACGGTGAGAACCTTGTGGGGGATGAC	1016
416870701_Ins1_C_Du	GTGGGCCGATATCTCTCCTAAGGTAAACGGTGAGAACCTTGTGGGGGATGAC	976
	<div>W A</div>	
416870701_Ins1_D_Du	AGATTTTGATATCTCTCCTAAGGTAAACGGTGAGAACCTTGTGGGGGATGAC	980
416870701_Ins1_E_Du	AGATTTTGATATCTCTCCTAAGGTAAACGGTGAGAACCTTGTGGGGGATGAC	988
Score	*****	

Appendix 10: Alignment of pACYC-F1::2 sequencing results.

Alignment of pACYC-F1::2 constructs using DNAdynamo. Cyan box indicates residues in WT and red box indicates the changed residues in mutant. From the alignment, only F1::2-A, -B and -C from Stellar cells confirmed with loop mutant.



Appendix 11: Alignment of pACYC-F1::3 sequencing results.

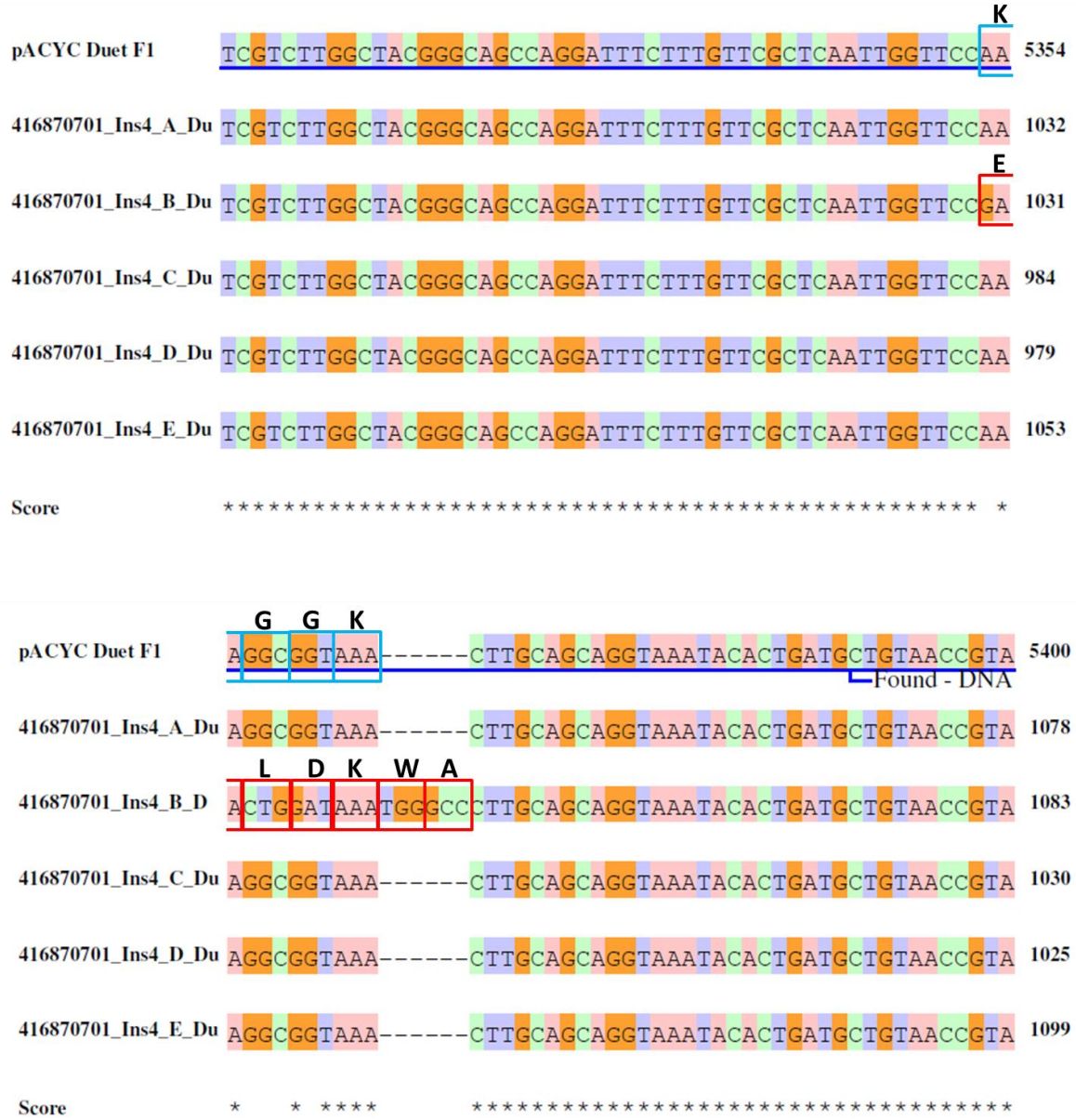
Alignment of pACYC-F1::3 constructs using DNAdynamo. Cyan box indicates residues in WT and red box indicates the changed residues in mutant. From the alignment, only F1::3-E, -G and -J from Stellar cells confirmed with loop mutant.

pACYC Duet F1	CCGCATCACTCTTACATATAAGGAAGGCGCTCCAATTACAATTATGGAC---	5067
416417601_F1_3_E_Du	CCGCATCACTCTTACATATAAGGAAGGCGCTCCAATTACAATTATGGAACTG	707
416417601_F1_3_G_Du	CCGCATCACTCTTACATATAAGGAAGGCGCTCCAATTACAATTATGGAACTG	736
416417601_F1_3_H_Du	CCGCATCACTCTTACATATAAGGAAGGCGCTCCAATTACAATTATGGAC---	768
416417601_F1_3_J_Du	CCGCATCACTCTTACATATAAGGAAGGCGCTCCAATTACAATTATGGAACTG	776
Score	*****	

pACYC Duet F1	---AATGG---AAACATCGATACAGAATTACTTGTTGGTACGCTTACTC	5110
416417601_F1_3_E_Du	GATAAATGGGCGGAAACATCGATACAGAATTACTTGTTGGTACGCTTACTC	759
416417601_F1_3_G_Du	GATAAATGGGCGGAAACATCGATACAGAATTACTTGTTGGTACGCTTACTC	788
416417601_F1_3_H_Du	---AATGG---AAACATCGATACAGAATTACTTGTTGGTACGCTTACTC	811
416417601_F1_3_J_Du	GATAAATGGGCGGAAACATCGATACAGAATTACTTGTTGGTACGCTTACTC	828
Score	*****	

Appendix 12: Alignment of pACYC-F1::4 sequencing results.

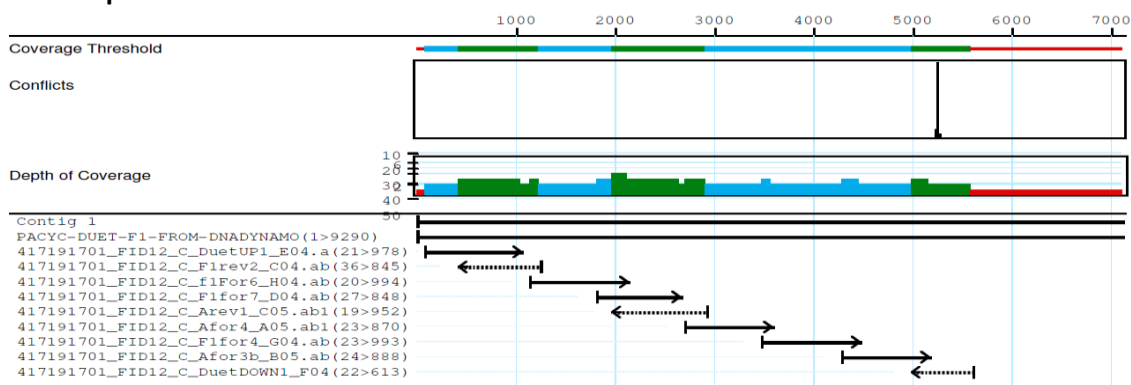
Alignment of pACYC-F1::4 constructs using DNAdynamo. Cyan box indicates residues in WT and red box indicates the changed residues in mutant. From the alignment, only F1::4-B from Stellar cells confirmed with loop mutant.



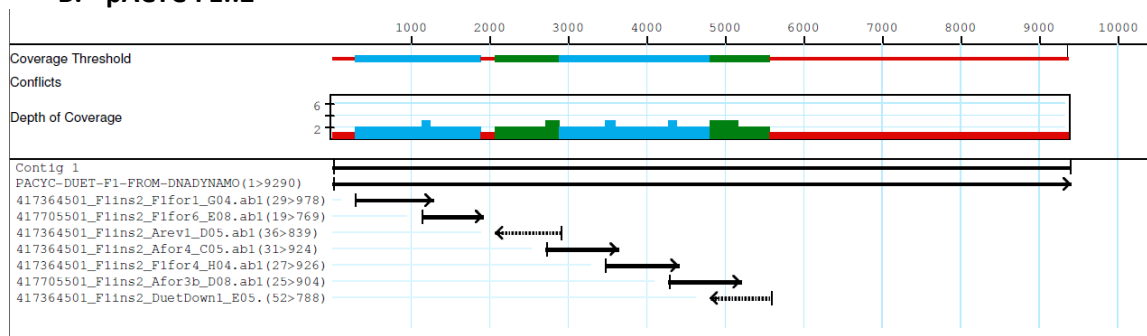
Appendix 13: Whole operon sequencing of pACYC-F1D12, F1::2, F1::3 and F1::4 constructs

Assembly of pACYC-F1D12, F1::2, F1::3 and F1::4 constructs in strategy view from DNASTar programme. From the assembly, all constructs were confirmed to have loop mutation and do not have any additional spontaneous mutation. In table are primers used for whole operon sequencing.

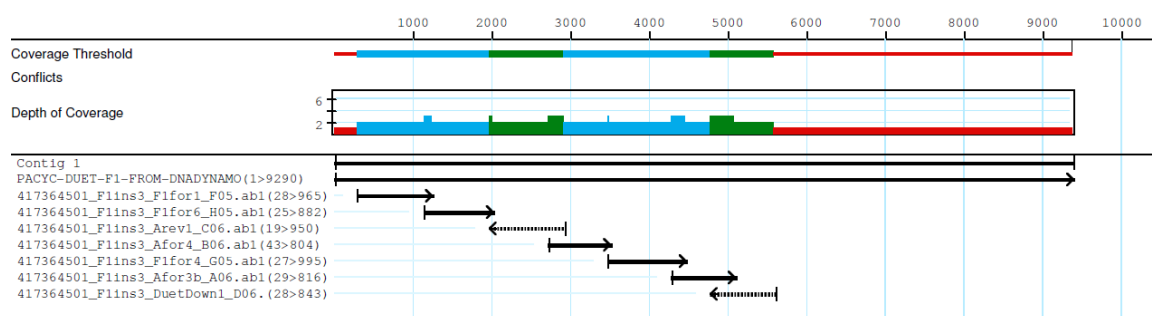
A. pACYC-F1D12



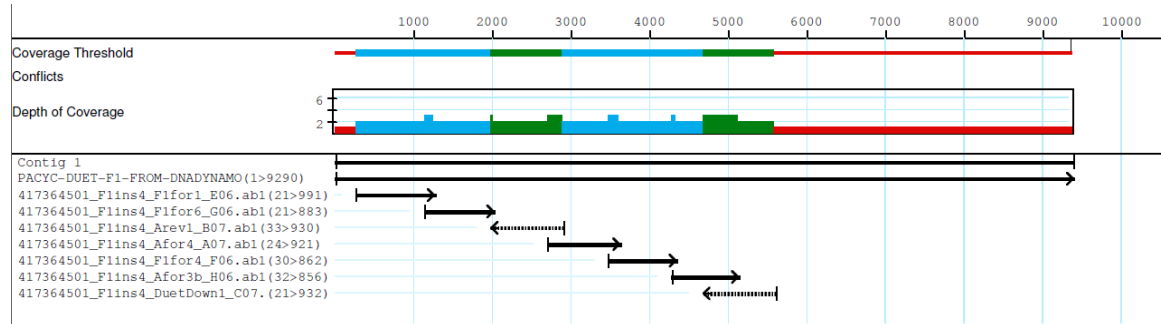
B. pACYC-F1::2



C. pACYC-F1::3



D. pACYC-F1::4



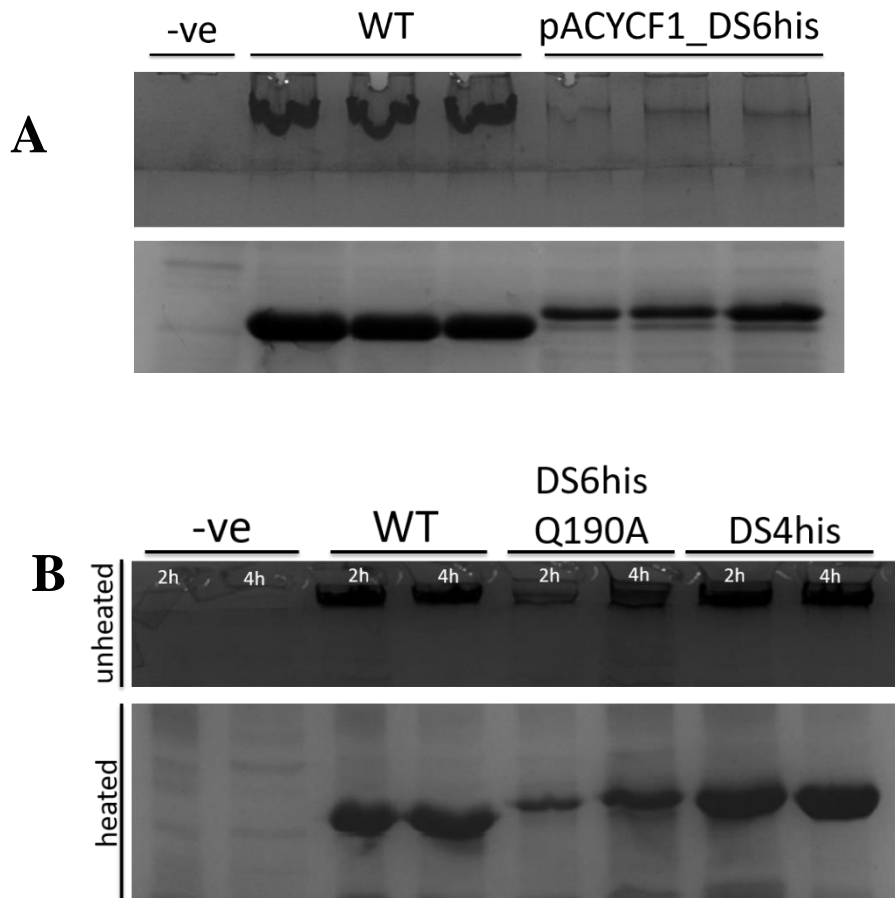
E. Sequencing primers of whole *caf* operon

Primer	Sequence 5'-3'	Priming site	Direction
DuetDOWN1	GAT TAT GCG GCC GTG TAC AA	Primes towards end of CafI	Reverse
DuetUP1	GTC CGG GAT CTC GAC GCT CTC CCT TAT G	Primes towards beginning of CafIR	Forward
F1#1	ATG TTG GGT CGA ACA TAA ATC G	Primes towards beginning of CafIR	Forward
F1#4	CAC AAA CAT CGT ATG TTG GCG CTC	Primes from middle of CafIA	Forward
F1#6	CCT TCT GCT GAA TCC TGA ATA C	Primes from end of CafIR towards intergenic promoter region	Forward
F1-7for	TCA GAG GAT CCT TTC GTG GTC AC	Primes from middle of CafIM towards	Forward
F1rev2	GAT TGG TGT TGC TGG GTA TG	Primes from start of intergenic promoter region towards CafIR	Reverse
CafIArev-1	CTA CGA AAG CGC CAA GCC CCT ATG	Primes from quarter of CafIA towards CafIM	Reverse
CafIAfor-3b	GAG GAA CTT ATG GGG AGA TCA GTG	Primes from end of CafIA towards CafI	Forward
CafIAfor-4	AAA TCT GGT ACA GAG CAA TGT G	Primes from start of CafIA	Forward

Appendix 14: Model of protein in this study attached in a CD

Name of Protein Model	Name of file in CD	Figure in thesis
Caf1 ₂ :Caf1M complex (PDB: 1Z9S) showing surface loops tested for epitope replacement in previous study	Epitope labeled.pse	Figure 5.3
Translocon of WT (Caf1A model based on FimD accommodate FimH)	Wildtype translocon.pse	Figure 4.6
Translocon of Caf1-F1D11	Translocon_F1D11.pse	Figure 5.14 (F1D11) (A)
Translocon of Caf1-F1::3	Translocon_F1ins3.pse	Figure 5.14 (F1::3) (B)
Translocon of Caf1-F1D12	Translocon_F1D12.pse	Figure 5.14 (F1D12) (A)
Translocon of Caf1-F1::2	Translocon_F1ins2.pse	Figure 5.14 (F1::2) (B)
Translocon of Caf1-F1D13	Translocon_F1D13.pse	Figure 5.14 (F1D13)v(A)
Translocon of Caf1-F1::1	Translocon_F1ins1.pse	Figure 5.14 (F1::1) (B)
Translocon of Caf1-F1D14	Translocon_F1D14.pse	Figure 5.14 (F1D14) (A)
Translocon of Caf1-F1::4	Translocon_F1::4.pse	Figure 5.14 (F1::4) (B)
Caf1A barrel, residues facing in labelled	Caf1A_barrel_residues_facing_in.pse	Figure 4.10 (B)
FimD barrel, residues facing in labelled	Fim_barrel_residues_facing_in.pse	Figure 4.10 (B)
<i>E. Coli</i> SE11 barrel, residues facing in labelled	SE11_barrel_residues_facing_in.pse	Figure 4.10 (B)
Closed Caf1A model and interactions labelled	Closed CAf1A labelled.pse	Figure 4.8 (C)
Superpose of closed PapC and closed Caf1A model	PapC_CAf1A_superpose_interactions.pse	Figure 4.12 and Figure 4.13

Appendix 15: SDS-PAGE of F1 from cells carrying polyhistidine insertion at loop 5



SDS-PAGE of surface F1 containing polyhistidine inserted at loop 5. (A) Showing heated and unheated triplicate sample of pACYCF1_DS6his. (B) Showing heated and unheated samples of pACYCF1_DS6his_Q169A and pACYCF1_DS4his. F1 was extracted at 65°C for 1 hour and run on SDS-PAGE showing unheated sample with high molecular weight polymer in stacking gel and monomer of Caf1 in separating gel. The heated sample was mixed with SDS loading dye and heated at >95°C for 15 minutes to denature all F1 fibre to Caf1 monomer. 3 OD cells after 4 hour thermoinduction were recovered, extracted and resuspend in 100µl PBS.

References

- Achtman, M., Zurth, K., Morelli, G., Torrea, G., Guiyoule, A. & Carniel, E. (1999). *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci U S A*, **96**, 14043-8.
- Anderson, G. W., Jr., Leary, S. E., Williamson, E. D., Titball, R. W., Welkos, S. L., Worsham, P. L. & Friedlander, A. M. (1996). Recombinant V antigen protects mice against pneumonic and bubonic plague caused by F1-capsule-positive and -negative strains of *Yersinia pestis*. *Infect Immun*, **64**, 4580-5.
- Andrews, G. P., Heath, D. G., Anderson, G. W., Jr., Welkos, S. L. & Friedlander, A. M. (1996). Fraction 1 capsular antigen (F1) purification from *Yersinia pestis* CO92 and from an *Escherichia coli* recombinant strain and efficacy against lethal plague challenge. *Infect Immun*, **64**, 2180-7.
- Atlas, R. M. (1998). The medical threat of biological weapons. *Crit Rev Microbiol*, **24**, 157-68.
- Baga, M., Norgren, M. & Normark, S. (1987). Biogenesis of *E. coli* Pap pili: papH, a minor pilin subunit involved in cell anchoring and length modulation. *Cell*, **49**, 241-51.
- Bendiner, E. (1989). Alexandre Yersin: pursuer of plague. *Hosp Pract (Off Ed)*, **24**, 121-8, 131-2, 135-8 passim.
- Bos, K. I., Schuenemann, V. J., Golding, G. B., Burbano, H. A., Waglechner, N., Coombes, B. K., McPhee, J. B., DeWitte, S. N., Meyer, M., Schmedes, S., Wood, J., Earn, D. J., Herring, D. A., Bauer, P., Poinar, H. N. & Krause, J. (2011). A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature*, **478**, 506-10.
- Boulanger, L. L., Ettestad, P., Fogarty, J. D., Dennis, D. T., Romig, D. & Mertz, G. (2004). Gentamicin and tetracyclines for the treatment of human plague: review of 75 cases in new Mexico, 1985-1999. *Clin Infect Dis*, **38**, 663-9.

- Brubaker, R. R. (1972). The genus *Yersinia*: biochemistry and genetics of virulence. *Curr Top Microbiol Immunol*, **57**, 111-58.
- Brubaker, R. R. (2003). Interleukin-10 and inhibition of innate immunity to *Yersinia*: roles of Yops and LcrV (V antigen). *Infect Immun*, **71**, 3673-81.
- Bullitt, E. & Makowski, L. (1995). Structural polymorphism of bacterial adhesion pili. *Nature*, **373**, 164-7.
- Busch, A. & Waksman, G. (2012). Chaperone-ushe pathways: diversity and pilus assembly mechanism. *Philos Trans R Soc Lond B Biol Sci*, **367**, 1112-22.
- Butler, T. (2009). Plague into the 21st century. *Clin Infect Dis*, **49**, 736-42.
- Chanteau, S., Rabarijaona, L., O'Brien, T., Rahalison, L., Hager, J., Boisier, P., Burans, J. & Rasolomaharo, M. (1998). F1 antigenaemia in bubonic plague patients, a marker of gravity and efficacy of therapy. *Trans R Soc Trop Med Hyg*, **92**, 572-3.
- Chanteau, S., Rahalison, L., Ralafiarisoa, L., Foulon, J., Ratsitorahina, M., Ratsifasoamanana, L., Carniel, E. & Nato, F. (2003). Development and testing of a rapid diagnostic test for bubonic and pneumonic plague. *Lancet*, **361**, 211-6.
- Chanteau, S., Rahalison, L., Ratsitorahina, M., Mahafaly, Rasolomaharo, M., Boisier, P., O'Brien, T., Aldrich, J., Keleher, A., Morgan, C. & Burans, J. (2000). Early diagnosis of bubonic plague using F1 antigen capture ELISA assay and rapid immunogold dipstick. *Int J Med Microbiol*, **290**, 279-83.
- Chapman, D. A., Zavialov, A. V., Chernovskaya, T. V., Karlyshev, A. V., Zav'yalova, G. A., Vasiliev, A. M., Dudich, I. V., Abramov, V. M., Zav'yalov, V. P. & MacIntyre, S. (1999). Structural and functional significance of the FGL sequence of the periplasmic chaperone Caf1M of *Yersinia pestis*. *J Bacteriol*, **181**, 2422-9.
- Cherepanov, P. A., Mikhailova, T. G., Karimova, G. A., Zakharova, N. M., Ershov Iu, V. & Volkovoi, K. I. (1991). [Cloning and detailed mapping of the fra-ymt region of the *Yersinia pestis* pFra plasmid]. *Mol Gen Mikrobiol Virusol*, 19-26.
- Choudhury, D., Thompson, A., Stojanoff, V., Langermann, S., Pinkner, J., Hultgren, S. J. & Knight, S. D. (1999). X-ray structure of the FimC-FimH chaperone-adhesin complex from uropathogenic *Escherichia coli*. *Science*, **285**, 1061-6.
- Cohen, R. J. & Stockard, J. L. (1967). Pneumonic plague in an untreated plague-vaccinated individual. *JAMA*, **202**, 365-6.

- Cornelius, C. A., Quenee, L. E., Overheim, K. A., Koster, F., Brasel, T. L., Elli, D., Ciletti, N. A. & Schneewind, O. (2008). Immunization with recombinant V10 protects cynomolgus macaques from lethal pneumonic plague. *Infect Immun*, **76**, 5588-97.
- Darby, C., Hsu, J. W., Ghorri, N. & Falkow, S. (2002). *Caenorhabditis elegans*: plague bacteria biofilm blocks food intake. *Nature*, **417**, 243-4.
- Darling, R. G., Catlett, C. L., Huebner, K. D. & Jarrett, D. G. (2002). Threats in bioterrorism. I: CDC category A agents. *Emerg Med Clin North Am*, **20**, 273-309.
- Deng, W., Burland, V., Plunkett, G., 3rd, Boutin, A., Mayhew, G. F., Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C., Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plano, G. V., Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S., Blattner, F. R. & Perry, R. D. (2002). Genome sequence of *Yersinia pestis* KIM. *J Bacteriol*, **184**, 4601-11.
- Derbise, A., Cerda Marin, A., Ave, P., Blisnick, T., Huerre, M., Carniel, E. & Demeure, C. E. (2012). An encapsulated *Yersinia pseudotuberculosis* is a highly efficient vaccine against pneumonic plague. *PLoS Negl Trop Dis*, **6**, e1528.
- Di Yu, X., Dubnovitsky, A., Pudney, A. F., Macintyre, S., Knight, S. D. & Zavialov, A. V. (2012). Allosteric mechanism controls traffic in the chaperone/usher pathway. *Structure*, **20**, 1861-71.
- Dodson, K. W., Jacob-Dubuisson, F., Striker, R. T. & Hultgren, S. J. (1993). Outer-membrane PapC molecular usher discriminately recognizes periplasmic chaperone-pilus subunit complexes. *Proc Natl Acad Sci U S A*, **90**, 3670-4.
- Du, Y., Rosqvist, R. & Forsberg, A. (2002). Role of fraction 1 antigen of *Yersinia pestis* in inhibition of phagocytosis. *Infect Immun*, **70**, 1453-60.
- Dubnovitsky, A. P., Duck, Z., Kersley, J. E., Hard, T., MacIntyre, S. & Knight, S. D. (2010). Conserved hydrophobic clusters on the surface of the Caf1A usher C-terminal domain are important for F1 antigen assembly. *J Mol Biol*, **403**, 243-59.
- Eisen, R. J., Bearden, S. W., Wilder, A. P., Montenieri, J. A., Antolin, M. F. & Gage, K. L. (2006). Early-phase transmission of *Yersinia pestis* by unblocked fleas as a mechanism explaining rapidly spreading plague epizootics. *Proc Natl Acad Sci U S A*, **103**, 15380-5.
- Fellows, P., Adamovicz, J., Hartings, J., Sherwood, R., Mega, W., Brasel, T., Barr, E., Holland, L., Lin, W., Rom, A., Blackwelder, W., Price, J., Morris, S., Snow, D. & Hart, M. K. (2010). Protection in mice passively immunized with serum

- from cynomolgus macaques and humans vaccinated with recombinant plague vaccine (rF1V). *Vaccine*, **28**, 7748-56.
- Feodorova, V. A. & Corbel, M. J. (2009). Prospects for new plague vaccines. *Expert Rev Vaccines*, **8**, 1721-38.
- Feodorova, V. A. & Motin, V. L. (2012). Plague vaccines: current developments and future perspectives. *Emerg Microbes Infect*, **1**, e36.
- Ford, B., Rego, A. T., Ragan, T. J., Pinkner, J., Dodson, K., Driscoll, P. C., Hultgren, S. & Waksman, G. (2010). Structural homology between the C-terminal domain of the PapC usher and its plug. *J Bacteriol*, **192**, 1824-31.
- Forman, S., Wulff, C. R., Myers-Morales, T., Cowan, C., Perry, R. D. & Straley, S. C. (2008). yadBC of *Yersinia pestis*, a new virulence determinant for bubonic plague. *Infect Immun*, **76**, 578-87.
- Fronzes, R., Remaut, H. & Waksman, G. (2008). Architectures and biogenesis of non-flagellar protein appendages in Gram-negative bacteria. *EMBO J*, **27**, 2271-80.
- Galyov, E. E., Smirnov, O., Karlishev, A. V., Volkovoy, K. I., Denesyuk, A. I., Nazimov, I. V., Rubtsov, K. S., Abramov, V. M., Dalvadyanz, S. M. & Zav'yalov, V. P. (1990). Nucleotide sequence of the *Yersinia pestis* gene encoding F1 antigen and the primary structure of the protein. Putative T and B cell epitopes. *FEBS Lett*, **277**, 230-2.
- Geibel, S., Procko, E., Hultgren, S. J., Baker, D. & Waksman, G. (2013). Structural and energetic basis of folded-protein transport by the FimD usher. *Nature*, **496**, 243-6.
- Gilleland, H. E., Jr., Gilleland, L. B. & Matthews-Greer, J. M. (1988). Outer membrane protein F preparation of *Pseudomonas aeruginosa* as a vaccine against chronic pulmonary infection with heterologous immunotype strains in a rat model. *Infect Immun*, **56**, 1017-22.
- Hahn, E., Wild, P., Hermanns, U., Sebbel, P., Glockshuber, R., Haner, M., Taschner, N., Burkhard, P., Aebi, U. & Muller, S. A. (2002). Exploring the 3D molecular architecture of *Escherichia coli* type 1 pili. *J Mol Biol*, **323**, 845-57.
- Hatkoff, M., Runco, L. M., Pujol, C., Jayatilaka, I., Furie, M. B., Bliska, J. B. & Thanassi, D. G. (2012). Roles of chaperone/usher pathways of *Yersinia pestis* in a murine model of plague and adhesion to host cells. *Infect Immun*, **80**, 3490-500.
- Henderson, N. S., Ng, T. W., Talukder, I. & Thanassi, D. G. (2011). Function of the usher N-terminus in catalysing pilus assembly. *Mol Microbiol*, **79**, 954-67.

- Hinnebusch, B. J. (2005). The evolution of flea-borne transmission in *Yersinia pestis*. *Curr Issues Mol Biol*, **7**, 197-212.
- Hinnebusch, B. J., Rudolph, A. E., Cherepanov, P., Dixon, J. E., Schwan, T. G. & Forsberg, A. (2002). Role of *Yersinia murine* toxin in survival of *Yersinia pestis* in the midgut of the flea vector. *Science*, **296**, 733-5.
- Huang, X. Z. & Lindler, L. E. (2004). The pH 6 antigen is an antiphagocytic factor produced by *Yersinia pestis* independent of *Yersinia* outer proteins and capsule antigen. *Infect Immun*, **72**, 7212-9.
- Huang, Y., Smith, B. S., Chen, L. X., Baxter, R. H. & Deisenhofer, J. (2009). Insights into pilus assembly and secretion from the structure and functional characterization of usher PapC. *Proc Natl Acad Sci U S A*, **106**, 7403-7.
- Hung, D. L., Knight, S. D., Woods, R. M., Pinkner, J. S. & Hultgren, S. J. (1996). Molecular basis of two subfamilies of immunoglobulin-like chaperones. *EMBO J*, **15**, 3792-805.
- Infectious Diseases Society of America., Memorial Institute for Infectious Diseases (Chicago Ill.), John McCormick Institute for Infectious Diseases. & John Rockefeller McCormick Memorial Fund. The journal of infectious diseases : official publication of the Infectious Diseases Society of America. Chicago, Ill.
- Cary, NC: University of Chicago Press
- Oxford University Press.
- Inglesby, T. V., Dennis, D. T., Henderson, D. A., Bartlett, J. G., Ascher, M. S., Eitzen, E., Fine, A. D., Friedlander, A. M., Hauer, J., Koerner, J. F., Layton, M., McDade, J., Osterholm, M. T., O'Toole, T., Parker, G., Perl, T. M., Russell, P. K., Schoch-Spana, M. & Tonat, K. (2000). Plague as a biological weapon: medical and public health management. Working Group on Civilian Biodefense. *JAMA*, **283**, 2281-90.
- Jacob-Dubuisson, F., Heuser, J., Dodson, K., Normark, S. & Hultgren, S. (1993). Initiation of assembly and association of the structural elements of a bacterial pilus depend on two specialized tip proteins. *EMBO J*, **12**, 837-47.
- Jones, C. H., Pinkner, J. S., Roth, R., Heuser, J., Nicholes, A. V., Abraham, S. N. & Hultgren, S. J. (1995). FimH adhesin of type 1 pili is assembled into a fibrillar tip structure in the Enterobacteriaceae. *Proc Natl Acad Sci U S A*, **92**, 2081-5.
- Jones, S. M., Griffin, K. F., Hodgson, I. & Williamson, E. D. (2003). Protective efficacy of a fully recombinant plague vaccine in the guinea pig. *Vaccine*, **21**, 3912-8.

- Keim, P. S. & Wagner, D. M. (2009). Humans and evolutionary and ecological forces shaped the phylogeography of recently emerged diseases. *Nat Rev Microbiol*, **7**, 813-21.
- Kersley, J. E., Zavialov, A. V., Moslehi, E., Knight, S. D. & MacIntyre, S. (2003). Mutagenesis elucidates the assembly pathway and structure of *Yersinia pestis* F1 polymer. *Adv Exp Med Biol*, **529**, 113-6.
- Knight, S. D. (2007). Structure and assembly of *Yersinia pestis* F1 antigen. *Adv Exp Med Biol*, **603**, 74-87.
- Knirel, Y. A., Dentovskaya, S. V., Senchenkova, S. N., Shaikhutdinova, R. Z., Kocharova, N. A. & Anisimov, A. P. (2006). Structural features and structural variability of the lipopolysaccharide of *Yersinia pestis*, the cause of plague. *J Endotoxin Res*, **12**, 3-9.
- Korea, C. G., Ghigo, J. M. & Beloin, C. (2011). The sweet connection: Solving the riddle of multiple sugar-binding fimbrial adhesins in *Escherichia coli*: Multiple *E. coli* fimbriae form a versatile arsenal of sugar-binding lectins potentially involved in surface-colonisation and tissue tropism. *Bioessays*, **33**, 300-11.
- Kringelum, J. V., Lundegaard, C., Lund, O. & Nielsen, M. (2012). Reliable B cell epitope predictions: impacts of method development and improved benchmarking. *PLoS Comput Biol*, **8**, e1002829.
- Kuehn, M. J., Heuser, J., Normark, S. & Hultgren, S. J. (1992). P pili in uropathogenic *E. coli* are composite fibres with distinct fibrillar adhesive tips. *Nature*, **356**, 252-5.
- Larsen, J. E., Lund, O. & Nielsen, M. (2006). Improved method for predicting linear B-cell epitopes. *Immunome Res*, **2**, 2.
- Lathem, W. W., Crosby, S. D., Miller, V. L. & Goldman, W. E. (2005). Progression of primary pneumonic plague: a mouse model of infection, pathology, and bacterial transcriptional activity. *Proc Natl Acad Sci U S A*, **102**, 17786-91.
- Le Trong, I., Aprikian, P., Kidd, B. A., Thomas, W. E., Sokurenko, E. V. & Stenkamp, R. E. (2010). Donor strand exchange and conformational changes during *E. coli* fimbrial formation. *J Struct Biol*, **172**, 380-8.
- Li, Q., Ng, T. W., Dodson, K. W., So, S. S., Bayle, K. M., Pinkner, J. S., Scarlata, S., Hultgren, S. J. & Thanassi, D. G. (2010). The differential affinity of the usher for chaperone-subunit complexes is required for assembly of complete pili. *Mol Microbiol*, **76**, 159-72.
- Ligon, B. L. (2006). Plague: a review of its history and potential as a biological weapon. *Semin Pediatr Infect Dis*, **17**, 161-70.

- Lindler, L. E., Plano, G. V., Burland, V., Mayhew, G. F. & Blattner, F. R. (1998). Complete DNA sequence and detailed analysis of the *Yersinia pestis* KIM5 plasmid encoding murine toxin and capsular antigen. *Infect Immun*, **66**, 5731-42.
- Loiez, C., Herwegh, S., Wallet, F., Armand, S., Guinet, F. & Courcol, R. J. (2003). Detection of *Yersinia pestis* in sputum by real-time PCR. *J Clin Microbiol*, **41**, 4873-5.
- MacIntyre, S., Zyrianova, I. M., Chernovskaya, T. V., Leonard, M., Rudenko, E. G., Zav'Yalov, V. P. & Chapman, D. A. (2001). An extended hydrophobic interactive surface of *Yersinia pestis* Caf1M chaperone is essential for subunit binding and F1 capsule assembly. *Mol Microbiol*, **39**, 12-25.
- Mapingire, O. S., Henderson, N. S., Duret, G., Thanassi, D. G. & Delcour, A. H. (2009). Modulating effects of the plug, helix, and N- and C-terminal domains on channel properties of the PapC usher. *J Biol Chem*, **284**, 36324-33.
- Meyer, K. F. (1970). Effectiveness of live or killed plague vaccines in man. *Bull World Health Organ*, **42**, 653-66.
- Meyer, K. F., Cavanaugh, D. C., Bartelloni, P. J. & Marshall, J. D., Jr. (1974a). Plague immunization. I. Past and present trends. *J Infect Dis*, **129**, Suppl:S13-8.
- Meyer, K. F., Smith, G., Foster, L., Brookman, M. & Sung, M. (1974b). Live, attenuated *Yersinia pestis* vaccine: virulent in nonhuman primates, harmless to guinea pigs. *J Infect Dis*, **129**, Suppl:S85-12.
- Mierzwa, J. (1975). The renal artery system in the dog. *Folia Morphol (Warsz)*, **34**, 337-48.
- Mizel, S. B., Graff, A. H., Sriranganathan, N., Ervin, S., Lees, C. J., Lively, M. O., Hantgan, R. R., Thomas, M. J., Wood, J. & Bell, B. (2009). Flagellin-F1-V fusion protein is an effective plague vaccine in mice and two species of nonhuman primates. *Clin Vaccine Immunol*, **16**, 21-8.
- Motin, V. L., Nakajima, R., Smirnov, G. B. & Brubaker, R. R. (1994). Passive immunity to yersiniae mediated by anti-recombinant V antigen and protein A-V antigen fusion peptide. *Infect Immun*, **62**, 4192-201.
- Mulvey, M. A., Lopez-Boado, Y. S., Wilson, C. L., Roth, R., Parks, W. C., Heuser, J. & Hultgren, S. J. (1998). Induction and evasion of host defenses by type 1-piliated uropathogenic *Escherichia coli*. *Science*, **282**, 1494-7.
- Mwengee, W., Butler, T., Mgema, S., Mhina, G., Almasi, Y., Bradley, C., Formanik, J. B. & Rochester, C. G. (2006). Treatment of plague with gentamicin or

- doxycycline in a randomized clinical trial in Tanzania. *Clin Infect Dis*, **42**, 614-21.
- Ng, T. W., Akman, L., Osisami, M. & Thanassi, D. G. (2004). The usher N terminus is the initial targeting site for chaperone-subunit complexes and participates in subsequent pilus biogenesis events. *J Bacteriol*, **186**, 5321-31.
- Nishiyama, M., Horst, R., Eidam, O., Herrmann, T., Ignatov, O., Vetsch, M., Bettendorff, P., Jelesarov, I., Grutter, M. G., Wuthrich, K., Glockshuber, R. & Capitani, G. (2005). Structural basis of chaperone-subunit complex recognition by the type 1 pilus assembly platform FimD. *EMBO J*, **24**, 2075-86.
- Nuccio, S. P. & Baumler, A. J. (2007). Evolution of the chaperone/usher assembly pathway: fimbrial classification goes Greek. *Microbiol Mol Biol Rev*, **71**, 551-75.
- Oshima, K., Toh, H., Ogura, Y., Sasamoto, H., Morita, H., Park, S. H., Ooka, T., Iyoda, S., Taylor, T. D., Hayashi, T., Itoh, K. & Hattori, M. (2008). Complete genome sequence and comparative analysis of the wild-type commensal *Escherichia coli* strain SE11 isolated from a healthy adult. *DNA Res*, **15**, 375-86.
- Oyston, P. C., Williamson, E. D., Leary, S. E., Eley, S. M., Griffin, K. F. & Titball, R. W. (1995). Immunization with live recombinant *Salmonella typhimurium* aroA producing F1 antigen protects against plague. *Infect Immun*, **63**, 563-8.
- Parkhill, J., Wren, B. W., Thomson, N. R., Titball, R. W., Holden, M. T., Prentice, M. B., Sebahia, M., James, K. D., Churcher, C., Mungall, K. L., Baker, S., Basham, D., Bentley, S. D., Brooks, K., Cerdeno-Tarraga, A. M., Chillingworth, T., Cronin, A., Davies, R. M., Davis, P., Dougan, G., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Karlyshev, A. V., Leather, S., Moule, S., Oyston, P. C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S. & Barrell, B. G. (2001). Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature*, **413**, 523-7.
- Perry, R. D. & Fetherston, J. D. (1997). *Yersinia pestis*--etiologic agent of plague. *Clin Microbiol Rev*, **10**, 35-66.
- Phan, G., Remaut, H., Wang, T., Allen, W. J., Pirker, K. F., Lebedev, A., Henderson, N. S., Geibel, S., Volkan, E., Yan, J., Kunze, M. B., Pinkner, J. S., Ford, B., Kay, C. W., Li, H., Hultgren, S. J., Thanassi, D. G. & Waksman, G. (2011). Crystal structure of the FimD usher bound to its cognate FimC-FimH substrate. *Nature*, **474**, 49-53.
- Prentice, M. B., James, K. D., Parkhill, J., Baker, S. G., Stevens, K., Simmonds, M. N., Mungall, K. L., Churcher, C., Oyston, P. C., Titball, R. W., Wren, B. W., Wain, J., Pickard, D., Hien, T. T., Farrar, J. J. & Dougan, G. (2001). *Yersinia*

- pestis pFra shows biovar-specific differences and recent common ancestry with a *Salmonella enterica* serovar Typhi plasmid. *J Bacteriol*, **183**, 2586-94.
- Prentice, M. B. & Rahalison, L. (2007). Plague. *Lancet*, **369**, 1196-207.
- Protsenko, O. A., Anisimov, P. I., Mozharov, O. T., Konnov, N. P. & Popov Iu, A. (1983). [Detection and characterization of the plasmids of the plague microbe which determine the synthesis of pesticin I, fraction I antigen and "mouse" toxin exotoxin]. *Genetika*, **19**, 1081-90.
- Rahalison, L., Vololonirina, E., Ratsitorahina, M. & Chanteau, S. (2000). Diagnosis of bubonic plague by PCR in Madagascar under field conditions. *J Clin Microbiol*, **38**, 260-3.
- Raoult, D., Mouffok, N., Bitam, I., Piarroux, R. & Drancourt, M. (2013). Plague: history and contemporary analysis. *J Infect*, **66**, 18-26.
- Rasmussen, S., Allentoft, M. E., Nielsen, K., Orlando, L., Sikora, M., Sjogren, K. G., Pedersen, A. G., Schubert, M., Van Dam, A., Kapel, C. M., Nielsen, H. B., Brunak, S., Avetisyan, P., Epimakhov, A., Khalyapin, M. V., Gnuni, A., Kriiska, A., Lasak, I., Metspalu, M., Moiseyev, V., Gromov, A., Pokutta, D., Saag, L., Varul, L., Yepiskoposyan, L., Sicheritz-Ponten, T., Foley, R. A., Lahr, M. M., Nielsen, R., Kristiansen, K. & Willerslev, E. (2015). Early divergent strains of *Yersinia pestis* in Eurasia 5,000 years ago. *Cell*, **163**, 571-82.
- Remaut, H., Rose, R. J., Hannan, T. J., Hultgren, S. J., Radford, S. E., Ashcroft, A. E. & Waksman, G. (2006). Donor-strand exchange in chaperone-assisted pilus assembly proceeds through a concerted beta strand displacement mechanism. *Mol Cell*, **22**, 831-42.
- Remaut, H., Tang, C., Henderson, N. S., Pinkner, J. S., Wang, T., Hultgren, S. J., Thanassi, D. G., Waksman, G. & Li, H. (2008). Fiber formation across the bacterial outer membrane by the chaperone/usher pathway. *Cell*, **133**, 640-52.
- Rollins, S. E., Rollins, S. M. & Ryan, E. T. (2003). *Yersinia pestis* and the plague. *Am J Clin Pathol*, **119 Suppl**, S78-85.
- Rosenzweig, J. A., Jejelowo, O., Sha, J., Erova, T. E., Brackman, S. M., Kirtley, M. L., van Lier, C. J. & Chopra, A. K. (2011). Progress on plague vaccine development. *Appl Microbiol Biotechnol*, **91**, 265-86.
- Russell, P., Eley, S. M., Hibbs, S. E., Manchee, R. J., Stagg, A. J. & Titball, R. W. (1995). A comparison of Plague vaccine, USP and EV76 vaccine induced protection against *Yersinia pestis* in a murine model. *Vaccine*, **13**, 1551-6.

- Russell, P. W. & Orndorff, P. E. (1992). Lesions in two *Escherichia coli* type 1 pilus genes alter pilus number and length without affecting receptor binding. *J Bacteriol*, **174**, 5923-35.
- Sauer, F. G., Futterer, K., Pinkner, J. S., Dodson, K. W., Hultgren, S. J. & Waksman, G. (1999). Structural basis of chaperone function and pilus biogenesis. *Science*, **285**, 1058-61.
- Sauer, F. G., Pinkner, J. S., Waksman, G. & Hultgren, S. J. (2002). Chaperone priming of pilus subunits facilitates a topological transition that drives fiber formation. *Cell*, **111**, 543-51.
- Saulino, E. T., Bullitt, E. & Hultgren, S. J. (2000). Snapshots of usher-mediated protein secretion and ordered pilus assembly. *Proc Natl Acad Sci U S A*, **97**, 9240-5.
- Saulino, E. T., Thanassi, D. G., Pinkner, J. S. & Hultgren, S. J. (1998). Ramifications of kinetic partitioning on usher-mediated pilus biogenesis. *EMBO J*, **17**, 2177-85.
- Sha, J., Agar, S. L., Baze, W. B., Olano, J. P., Fadl, A. A., Erova, T. E., Wang, S., Foltz, S. M., Suarez, G., Motin, V. L., Chauhan, S., Klimpel, G. R., Peterson, J. W. & Chopra, A. K. (2008). Braun lipoprotein (Lpp) contributes to virulence of yersiniae: potential role of Lpp in inducing bubonic and pneumonic plague. *Infect Immun*, **76**, 1390-409.
- Skurnik, M., Peippo, A. & Ervela, E. (2000). Characterization of the O-antigen gene clusters of *Yersinia pseudotuberculosis* and the cryptic O-antigen gene cluster of *Yersinia pestis* shows that the plague bacillus is most closely related to and has evolved from *Y. pseudotuberculosis* serotype O:1b. *Mol Microbiol*, **37**, 316-30.
- Slack, P. (1989). The black death past and present. 2. Some historical problems. *Trans R Soc Trop Med Hyg*, **83**, 461-3.
- Sokurenko, E. V., Courtney, H. S., Ohman, D. E., Klemm, P. & Hasty, D. L. (1994). FimH family of type 1 fimbrial adhesins: functional heterogeneity due to minor sequence variations among fimH genes. *J Bacteriol*, **176**, 748-55.
- Stacy, S., Pasquali, A., Sexton, V. L., Cantwell, A. M., Kraig, E. & Dube, P. H. (2008). An age-old paradigm challenged: old baboons generate vigorous humoral immune responses to LcrV, a plague antigen. *J Immunol*, **181**, 109-15.
- Stenseth, N. C., Atshabar, B. B., Begon, M., Belmain, S. R., Bertherat, E., Carniel, E., Gage, K. L., Leirs, H. & Rahalison, L. (2008). Plague: past, present, and future. *PLoS Med*, **5**, e3.

- Sun, W., Roland, K. L. & Curtiss, R., 3rd (2011). Developing live vaccines against plague. *J Infect Dev Ctries*, **5**, 614-27.
- Titball, R. W. & Williamson, E. D. (2001). Vaccination against bubonic and pneumonic plague. *Vaccine*, **19**, 4175-84.
- Titball, R. W. & Williamson, E. D. (2004). *Yersinia pestis* (plague) vaccines. *Expert Opin Biol Ther*, **4**, 965-73.
- Tomaso, H., Reisinger, E. C., Al Dahouk, S., Frangoulidis, D., Rakin, A., Landt, O. & Neubauer, H. (2003). Rapid detection of *Yersinia pestis* with multiplex real-time PCR assays using fluorescent hybridisation probes. *FEMS Immunol Med Microbiol*, **38**, 117-26.
- Verger, D., Miller, E., Remaut, H., Waksman, G. & Hultgren, S. (2006). Molecular mechanism of P pilus termination in uropathogenic *Escherichia coli*. *EMBO Rep*, **7**, 1228-32.
- Volkan, E., Kalas, V., Pinkner, J. S., Dodson, K. W., Henderson, N. S., Pham, T., Waksman, G., Delcour, A. H., Thanassi, D. G. & Hultgren, S. J. (2013). Molecular basis of usher pore gating in *Escherichia coli* pilus biogenesis. *Proc Natl Acad Sci U S A*, **110**, 20741-6.
- Williamson, E. D., Eley, S. M., Griffin, K. F., Green, M., Russell, P., Leary, S. E., Oyston, P. C., Easterbrook, T., Reddin, K. M., Robinson, A. & et al. (1995). A new improved sub-unit vaccine for plague: the basis of protection. *FEMS Immunol Med Microbiol*, **12**, 223-30.
- Williamson, E. D., Packer, P. J., Waters, E. L., Simpson, A. J., Dyer, D., Hartings, J., Twenhafel, N. & Pitt, M. L. (2011). Recombinant (F1+V) vaccine protects cynomolgus macaques against pneumonic plague. *Vaccine*, **29**, 4771-7.
- Wright, K. J., Seed, P. C. & Hultgren, S. J. (2007). Development of intracellular bacterial communities of uropathogenic *Escherichia coli* depends on type 1 pili. *Cell Microbiol*, **9**, 2230-41.
- Yu, X., Visweswaran, G. R., Duck, Z., Marupakula, S., MacIntyre, S., Knight, S. D. & Zavialov, A. V. (2009). Caf1A usher possesses a Caf1 subunit-like domain that is crucial for Caf1 fibre secretion. *Biochem J*, **418**, 541-51.
- Yu, X. D., Fooks, L. J., Moslehi-Mohebi, E., Tischenko, V. M., Askarieh, G., Knight, S. D., Macintyre, S. & Zavialov, A. V. (2012). Large is fast, small is tight: determinants of speed and affinity in subunit capture by a periplasmic chaperone. *J Mol Biol*, **417**, 294-308.
- Zav'yalov, V., Zavialov, A., Zav'yalova, G. & Korpela, T. (2010). Adhesive organelles of Gram-negative pathogens assembled with the classical

chaperone/usher machinery: structure and function from a clinical standpoint. *FEMS Microbiol Rev*, **34**, 317-78.

- Zav'yalov, V. P., Zav'yalova, G. A., Denesyuk, A. I. & Korpela, T. (1995). Modelling of steric structure of a periplasmic molecular chaperone Caf1M of *Yersinia pestis*, a prototype member of a subfamily with characteristic structural and functional features. *FEMS Immunol Med Microbiol*, **11**, 19-24.
- Zavialov, A., Zav'yalova, G., Korpela, T. & Zav'yalov, V. (2007). FGL chaperone-assembled fimbrial polyadhesins: anti-immune armament of Gram-negative bacterial pathogens. *FEMS Microbiol Rev*, **31**, 478-514.
- Zavialov, A. V., Berglund, J., Pudney, A. F., Fooks, L. J., Ibrahim, T. M., MacIntyre, S. & Knight, S. D. (2003). Structure and biogenesis of the capsular F1 antigen from *Yersinia pestis*: preserved folding energy drives fiber formation. *Cell*, **113**, 587-96.
- Zavialov, A. V., Kersley, J., Korpela, T., Zav'yalov, V. P., MacIntyre, S. & Knight, S. D. (2002). Donor strand complementation mechanism in the biogenesis of non-pilus systems. *Mol Microbiol*, **45**, 983-95.
- Zavialov, A. V., Tischenko, V. M., Fooks, L. J., Brandsdal, B. O., Aqvist, J., Zav'yalov, V. P., Macintyre, S. & Knight, S. D. (2005). Resolving the energy paradox of chaperone/usher-mediated fibre assembly. *Biochem J*, **389**, 685-94.
- Zhang, Y. & Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res*, **33**, 2302-9.
- Zhou, D. & Yang, R. (2009). Molecular Darwinian evolution of virulence in *Yersinia pestis*. *Infect Immun*, **77**, 2242-50.